

# Repositórios digitais teoria e prática

Fernando Vechiato  
Clediane Guedes  
Débora Koshiyama  
Elisângela Moura  
Emanuelle Torino  
Maria Aniolly Maia  
Tércia Marques  
(organizadores)







# Repositórios digitais teoria e prática





UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

*Reitor* Luiz Alberto Pilatti  
*Vice-Reitora* Vanessa Ishikawa Rasoto



EDITORA DA UTFPR

*Coordenadora-Geral* Camila Lopes Ferreira  
*Coordenadora-Adjunta* Emanuelle Torino

#### CONSELHO EDITORIAL


*Titulares* Anna Luiza Metidierl Cruz Malthez  
Awdry Feisser Miquelin  
Douglas Sampaio Henrique  
Eduardo Leite kruger  
Francis Kanashiro Meneghetti  
Ligia Patrícia Torino Guassu  
Marcos Antônio Florczak  
Rogério Caetano de Almeida  
Thomaz Aurélio Pagioro

*Suplentes* Adriane de Lima Penteado  
Alberto Yoshihiro Nakano  
Alessandra Dutra  
Anderson Catapan  
Cintia de Lourdes Nahhas Rodacki  
Ricardo Luders  
Ricardo Yuji Sado  
Rodrigo Alexandre de Carvalho Xavier  
Sara Tatiana Moreira

Fernando Vechiato  
Clediane Guedes  
Débora Koshiyama  
Elisângela Moura  
Emanuelle Torino  
Maria Aniolly Maia  
Tércia Marques  
(organizadores)

# Repositórios digitais teoria e prática

© 2017 Editora da Universidade Tecnológica Federal do Paraná

 Esta licença permite o download e o compartilhamento da obra desde que sejam atribuídos créditos ao(s) autor(es), sem a possibilidade de alterá-la ou utilizá-la para fins comerciais.

Disponível em: <<http://repositorio.utfpr.edu.br/jspui/>>.

---

#### Dados Internacionais de Catalogação na Publicação

---

R425r      Repositórios digitais : teoria e prática / organização de Fernando Vechiato ...  
[et al.]. – Curitiba: EDUTFPR, 2017.  
271 p.: il. ; 23 cm.  
ISBN: 978-85-7014-196-5 (Impresso)  
ISBN: 978-85-7014-197-2 (E-book)  
E-book disponível em: <http://repositorio.utfpr.edu.br/jspui/>  
1. Repositórios institucionais. 2. Sistemas de recuperação da informação. 3.  
Tecnologia - Serviços de informação. 4. Ciência da informação. I. Vechiato, Fernando  
Luiz, org. II. Título.

CDD (23. ed.) 020

---

Bibliotecário: Walison Oliveira CRB-9/1871

*Coordenação Editorial*      Camila Lopes Ferreira  
Emanuelle Torino

*Projeto Gráfico*      Vanessa Constance Ambrosio

*Revisão*      Adão de Araújo

#### **EDUTFPR**

Editora da Universidade Tecnológica Federal do Paraná  
Av. Sete de Setembro, 3165  
80230-901 Curitiba PR  
[www.utfpr.edu.br/editora](http://www.utfpr.edu.br/editora)

# SUMÁRIO

- 7**    PREFÁCIO
- 9**    APRESENTAÇÃO
- 13**   O VOO DA CRISÁLIDA  
Marcos Galindo
- 33**   GESTÃO INTEGRADA DA INFORMAÇÃO CIENTÍFICA E  
TECNOLÓGICA E O ACESSO ABERTO: ONDE ESTAMOS E ONDE  
PODEMOS CHEGAR  
Fernando César Lima Leite e Michelli Pereira da Costa
- 65**   TECNOLOGIAS PARA GESTÃO DA INFORMAÇÃO  
Milton Shintaku
- 91**   POLÍTICAS EM REPOSITÓRIOS DIGITAIS: DAS DIRETRIZES À  
IMPLEMENTAÇÃO  
Emanuelle Torino
- 115**   REPOSITÓRIO INSTITUCIONAL UNESP: INOVAÇÃO NO PROCESSO  
DE CRIAÇÃO  
Silvana Aparecida Borsetti Gregorio Vidotti, Flávia Maria Bastos e  
Felipe Augusto Arakaki
- 143**   O PAPEL DOS REPOSITÓRIOS DIGITAIS NA CURADORIA DE DADOS  
DE PESQUISA  
Luis Fernando Sayão
- 167**   ADOÇÃO DE PADRÕES DE METADADOS PARA REPOSITÓRIOS DE  
DADOS DIGITAIS NA CIÊNCIA ABERTA  
Adriana Carla Silva de Oliveira

- 193** DESAFIOS À COMUNIDADE IBERO-AMERICANA DE METADADOS  
EM REPOSITÓRIOS DIGITAIS PARA MAXIMIZAÇÃO DA  
INTEROPERABILIDADE  
Ana Alice Baptista
- 205** METADADOS E CURADORIA DIGITAL: UM PERFIL DE APLICAÇÃO  
PARA DESCRIÇÃO DE DADOS DE INVESTIGAÇÃO NA ÁREA DA  
OCEANOGRAFIA BIOLÓGICA  
Rúbia Tatiana Gattelli
- 227** GESTÃO DE DADOS MULTIDISCIPLINAR NA UNIVERSIDADE DO  
PORTO: EXPERIÊNCIAS EM GRUPOS DE DIMENSÕES DIVERSAS  
Cristina Ribeiro, João Rocha da Silva, João Aguiar Castro, Ricardo  
Carvalho Amorim, João Correia Lopes e Gabriel David
- 245** DSPACE CRIS  
Lucas Rodrigues da Costa, Milton Shintaku e Lucas Ângelo da  
Silveira
- 265** ORGANIZADORES
- 267** AUTORES



# PREFÁCIO

Refletir sobre Repositórios Digitais, na perspectiva teórico-prática, implica uma relação mútua de saberes para o processo de implementação desses repositórios. No caso específico desta publicação, isso se deu por meio da incorporação de conhecimento de seus diversos atores sociais e, ainda, de áreas distintas do saber, conforme explicitado nos capítulos que a compõem. Com efeito, essa relação é demonstrada por meio de reflexões apresentadas em toda a sua estrutura textual, a qual reúne produções e resultados originados no I Fórum Nacional de Repositórios Digitais (FNRD). Assim, numa abordagem de ordem teórico-conceitual, essas reflexões envolvem desde conceitos básicos relacionados ao tema, por exemplo, gestão, informação científica e tecnológica, tecnologias, acesso aberto, armazenamento, metadados, padrões, curadoria digital, preservação – até observações inerentes à relação pesquisa-prática na implementação desses repositórios.

Na atualidade, acredita-se que o quadro de referência sobre Repositórios Digitais, em nível internacional, decorre da convergência de três fatores: o destaque no fenômeno contemporâneo da informação, o avanço das tecnologias de informação e comunicação e, também, o aumento vertiginoso da produção do conhecimento e da informação científica e tecnológica. Isso tem proporcionado um novo ciclo de comunicação científica em todos os campos do conhecimento, haja vista a importância do acesso aberto no novo universo e, ainda, as novas formas de organização, acesso e recuperação da informação, notadamente, a científica e a tecnológica.

A partir dessa compreensão, percebe-se que os repositórios digitais, enquanto sistemas de acesso aberto, se constituem como um espaço privilegiado onde existe uma política que determina implementações que envolvem, necessariamente, ações-reflexões-ações para a qualificação dos seus metadados. Em consequência, esses metadados têm contribuído, de modo geral, para com os processos de difusão e de localização de dados de pesquisa, entre outros.

Nessa perspectiva, tornou-se notória a importância da discussão de base teórica sobre alguns conceitos apresentados neste livro e, que, ao mesmo tempo, circundam o trinômio: Repositórios Digitais – Acesso aberto – Tecnologias. Isso se justifica, tanto em função da importância da gestão da informação, por exemplo, a científica e a tecnológica, como a expansão da visibilidade de dados de pesquisa científica. Desse modo, a presente obra constitui, em um novo ciclo, um

diálogo com pesquisadores e profissionais interessados e comprometidos com a implementação de repositórios digitais, haja vista a possibilidade de uma gestão integrada da informação científica, tecnológica e técnica. Por isso, este livro traz contribuições de ordem teórico-prática, para quaisquer repositórios digitais.

Por fim, no desenvolvimento de cada aspecto específico sobre sua temática central, os capítulos que o compõem servem como contribuição para discussão acerca do tema. Nesse sentido, os autores ilustram uma ideia básica: repositórios digitais constituem, hoje, um papel essencial na atividade de comunicação científica. Assim, a proposta desta coletânea traz relevantes contribuições no sentido de explicitar a importância de aspectos teóricos e práticos inerentes a esses repositórios.

Rildecy Medeiros

# APRESENTAÇÃO

Os repositórios digitais são ambientes informacionais que se destacam no atual cenário científico e tecnológico por viabilizarem o armazenamento, a disseminação e a preservação: da produção intelectual, científica e/ou artística de uma instituição (repositórios institucionais), área do conhecimento (repositórios temáticos) ou mesmo de uma comunidade não necessariamente vinculada a uma instituição de ensino e pesquisa, no contexto do Acesso Aberto; e dos dados provenientes de pesquisas científicas (repositórios de dados de pesquisa), que podem ser compartilhados entre os pesquisadores de uma comunidade científica, visando seu uso e reuso no contexto da Ciência Aberta.

*Repositórios digitais: teoria e prática* é resultante de inquietações de pesquisadores e profissionais que observam a lacuna existente entre a pesquisa e a prática de implementação de repositórios digitais no cenário ibero-americano, inquietações estas que puderam ser discutidas por ocasião do I Fórum Nacional de Repositórios Digitais (FNRD), evento sediado pela Universidade Federal do Rio Grande do Norte (UFRN), em Natal/RN, no mês de novembro de 2016, e financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Esta obra dá continuidade e avança nas discussões realizadas no decorrer do referido evento, apontando para tendências teóricas e práticas na pesquisa e na implementação de repositórios digitais, o que decerto contribuirá para a comunidade acadêmico-científica e de prática ao abordar aspectos ligados à gestão e às políticas para repositórios digitais, aos metadados em repositórios digitais, às tecnologias para repositórios digitais e à preservação em repositórios digitais.

Os autores que assinam os onze capítulos desta obra são pesquisadores e profissionais que vêm dedicando seu trabalho ao Acesso Aberto e contribuem significativamente para os avanços dos estudos e práticas em repositórios digitais nos âmbitos científico, tecnológico e social.

*O voo da crisálida*, de Marcos Galindo, introduz a obra e elabora um resgate histórico-evolutivo em relação aos instrumentos de gestão do conhecimento sob a metáfora que empresta o título a este primeiro capítulo, visando observar e correlacionar eventos com o surgimento da curadoria digital.

Fernando César Lima Leite e Michelli Pereira da Costa assinam o segundo capítulo, intitulado *Gestão integrada da informação científica e tecnológica e o acesso aberto: onde estamos e onde podemos chegar*, e abordam a trajetória e as

perspectivas do acesso aberto no contexto brasileiro do ponto de vista da gestão da informação científica.

*Tecnologias para gestão da informação*, de Milton Shintaku, é o terceiro capítulo e apresenta tecnologias e ferramentas emergentes para a gestão da informação para diferentes finalidades e contextos, como bibliotecas, repositórios, eventos científicos entre outros.

Emanuelle Torino assina o quarto capítulo e aborda as *Políticas em repositórios digitais: das diretrizes à implementação*, detalhando aspectos que circundam as políticas de informação no contexto da gestão de repositórios digitais.

No quinto capítulo, intitulado *Repositório Institucional UNESP: inovação no processo de criação*, assinado por Silvana Aparecida Borsetti Gregorio Vidotti, Flávia Maria Bastos, Felipe Augusto Arakaki, é apresentada e detalhada a metodologia de coleta de dados realizada no âmbito do Repositório Nacional Institucional da UNESP, sendo esta considerada inovadora em comparação às metodologias geralmente implementadas.

Luís Fernando Sayão dedica o sexto capítulo, intitulado *O papel dos repositórios digitais na curadoria de dados de pesquisa*, visando lançar uma discussão pertinente para a comunicação científica a respeito do reuso dos dados de pesquisa no contexto da e-Science, enfatizando o papel dos repositórios digitais para a gestão desses dados.

*Adoção de padrões de metadados para repositórios de dados digitais na ciência aberta* é o sétimo capítulo deste livro, assinado por Adriana Carla Silva de Oliveira, que enfatiza a definição de padrões de metadados para repositórios de dados, considerando sua importância para a representação e para a recuperação de dados de pesquisa de diferentes domínios do conhecimento.

Ana Alice Baptista é autora do oitavo capítulo, intitulado *Desafios à comunidade Ibero-Americana de metadados em repositórios digitais para maximização da interoperabilidade*, em que apresenta uma proposta à referida comunidade baseada na criação de perfis de aplicação e ligação de dados, de modo a promover maior interoperabilidade entre os repositórios digitais de acesso aberto.

*Metadados e curadoria digital: um perfil de aplicação para descrição de dados de investigação na área de Oceanografia Biológica*, nono capítulo, assinado por Rúbia Tatiana Gattelli, apresenta o detalhamento do processo de construção de um perfil de aplicação de metadados para um domínio específico, neste caso, a área de Oceanografia Biológica.

O décimo capítulo é intitulado *Gestão de dados multidisciplinar na Universidade do Porto: experiências em grupos de dimensões diversas* e assinado pelos autores Cristina Ribeiro, João Rocha da Silva, João Aguiar Castro, Ricardo Carvalho

Amorim, João Correia Lopes, Gabriel David, em que tratam de iniciativas para a gestão de dados multidisciplinar com ênfase na plataforma Dendro.

O último capítulo é assinado por Lucas Rodrigues da Costa, Milton Shintaku, Lucas Ângelo da Silveira e intitulado *DSpace CRIS* e detalha a respeito desta tecnologia emergente que se caracteriza como uma extensão do *DSpace*, *software* livre utilizado amplamente para a gestão de repositórios institucionais brasileiros.

Desejamos a todos uma excelente leitura!

Os organizadores





# O VOO DA CRISÁLIDA

Marcos Galindo





*“[...] envolvermo-nos no casulo da nossa alma, fazermo-nos crisálida e aguardarmos a metamorfose, porque ela acaba sempre por chegar” (STRINDBERG apud MOURÃO, 2013, p. 63)*

## INTRODUÇÃO

Uma visão evolutiva sobre a natureza e o *status* dos instrumentos de gestão do conhecimento registrado nos permite traçar uma árvore genealógica para este ferramental, detectar as tendências que prosperaram e entender as possibilidades futuras do campo com base na compreensão de seus antecedentes históricos.

Nesse contexto, apresenta-se com especial relevância o estudo das vigências sociais dos formatos que lograram êxito na função de **interfacear** o conhecimento, bem como das estratégias de gestão demandadas sob a influência destas interfaces. Um novo limiar tecnológico já se anuncia em nosso tempo histórico e certamente exigirá novas e mais complexas ferramentas para ordenação do espaço da inteligência compartilhada. A curadoria digital vem emergindo como uma dessas ferramentas.

O voo da crisálida é uma metáfora que pretende representar a condição de metamorfose que reencontramos de tempos em tempos no clímax dos ciclos históricos. Procura expressar no mundo das ideias um conjunto de fenômenos sociais conexos que ligam as atividades humanas de resgate, tratamento, organização, preservação e acesso ao conhecimento em seu estado registral, reconhecendo suas múltiplas fases ou vigências sociais.

É natural que uma visão escatológica se manifeste nos estertores de cada ciclo de desenvolvimento da humanidade. E a certo modo este senso comum está correto, cada novo evento de mudança traz consigo um pouco do final de uma forma de humanidade que é sobreposta por uma nova mentalidade e expressão de contemporaneidade.

Crisálida é o estágio de pupa no qual a borboleta em fase de larva se recolhe num casulo tecido em seda que enrijece com o seu amadurecimento. É compreensível que, para a lagarta enclausurada, seu esquite dourado prenuncie a própria morte. De fato, a lagarta tem de desaparecer para a borboleta emergir. Todavia, esta mudança radical – denominada pelos gregos de **metamorfose** – implica não na morte, mas na transformação da lagarta que rasteja no imago que pode voar.

A exemplo da documentação – que se desenvolveu no início do século XX em resposta às necessidades de organização da informação demandadas pela

sociedade industrial –, a curadoria digital é resultado de uma tendência contemporânea imposta pela nova ambiência da sociedade da informação e pela instalação social do ciberespaço.

## ENTRE O SER E O ESTAR: A NATUREZA E O STATUS DAS COISAS

Na obra *Política*, Aristóteles<sup>1</sup> estabeleceu a distinção fundamental entre a natureza das coisas e seu *status*, sendo este referente ao estado transitório daquela natureza que é visível e penetrável à compreensão da inteligência humana. O *status* é temporal enquanto a natureza refere-se à forma ao que os seres são por nascimento (ARISTÓTELES, 1997).

Ortega y Gasset, (2006) a certa medida, recuperava essa reflexão aristotélica em *Misión del bibliotecário*<sup>2</sup>. Neste texto o autor se referia à noção de **vigência social do conhecimento** aplicado ao especial universo dos bibliotecários. A história, segundo ele: “é, principalmente, a história do surgimento, desenvolvimento e desaparecimento das vigências sociais” (ORTEGA Y GASSET, 2006, p. 19).

Sobre esta natureza transitória, explicava que a decisão sobre a permanência ou finitude das coisas não cabia ao indivíduo, mas à coletividade: “É indiferente se a pessoa está ou não de acordo com elas: sua vigência não depende de que tu ou eu prestemos nossa aprovação” (ORTEGA Y GASSET, 2006, p. 19).

No universo material, o conceito de vigência diz respeito ao lapso de tempo para o qual é aceitável o uso ou consumo de um determinado produto. No mundo das normas e regras, o termo é utilizado para definir a particularidade ou estado daquilo que é vigente, da valência ou vigor que possuem as coisas enquanto geram resultados. No mundo dos instrumentos, aplica-se o conceito para definir aquelas ferramentas que possuem função social ativa, cuja existência permanece justificada por uma larga base de usuários. Pode-se dizer, portanto, que se trata da temporalidade antrópica das coisas enquanto seu valor social é sustentável.

Os instrumentos são uma forma de prótese da inteligência humana que emulam ou realizam, para o homem que a opera, tarefas impossíveis ao seu aparato biológico. Desse modo, a tecnologia, não obstante ser protética – exterior à natureza humana – artificial ao seu ser biológico, é paradoxalmente natural ao seu

1 Na obra *Política*, Aristóteles discute sobre se a escravidão distingue entre aqueles que são escravos porque a lei diz que eles são e aqueles que são escravos por natureza, dizendo que o inquérito depende de se há algum desses escravos naturais. Somente alguém tão diferente de outras pessoas como o corpo é da alma ou os animais são de seres humanos seria um escravo por natureza, Aristóteles conclui, todos os outros sendo escravos exclusivamente por lei ou convenção. Alguns estudiosos concluíram, portanto, que as qualificações para a escravidão natural impedem a existência de tal ser.

2 *Misión del bibliotecário*. Trata-se do discurso proferido por Ortega y Gasset por ocasião da abertura do Segundo Congresso Internacional de Bibliotecas e Bibliografia realizado em Madrid aos 20 de maio de 1935.

ser intelectual. A tecnologia testemunha a capacidade criativa, usada por nós para encontrar soluções novas para novos problemas.

Criada pela necessidade e pela oportunidade oferecidas pelas novas demandas, a função dos instrumentos pode ser longa, mantendo a vigência das ferramentas enquanto seu uso for uma necessidade social. A tecnologia, *stricto sensu*, é o arcabouço material utilizado para operacionalizar a solução lógica desenvolvida para fazer frente às necessidades apresentadas pela interação do homem com seu meio.

Somente os sistemas melhor adaptados sobrevivem, esta condição implica na assunção de uma das leis fundamentais do evolucionismo, aquela que afirma que *a mudança é o único estado permanente de evolução* (MEGGINSON, 1963, p. 3)<sup>3</sup>. Por essa razão, a tecnologia depende da função social, a envolve e historicamente tem vida curta. É evolutiva e adapta-se e se reconstrói permanentemente, buscado eficiência e produtividade para atender as necessidades sociais no tempo e no espaço.

Assim, a função social surgida com um graveto, como aquele utilizado pelo homem pré-histórico para executar uma pintura rupestre na parede de uma caverna, sobrevive e emerge 30 mil anos depois em uma caneta esferográfica. A tecnologia, a seu turno, adapta-se e evolui sobre o mesmo espírito tecnológico modificando suas estratégias e interfaces. Assim, o graveto desaparece cedendo lugar a um instrumento sensivelmente mais complexo, preciso, eficiente e produtivo, sem perder, contudo, sua alma funcional.

Assim como os instrumentos, as profissões guardam estreita relação entre sua vigência e a função social que as orienta. Segundo Ortega y Gasset (2006, p. 16), o trabalho do profissional bibliotecário variou sempre “em função, rigorosamente, do que o livro significou como necessidade social”. Com base nessa reflexão, o autor postula que, para determinar a missão do bibliotecário, seria necessário, então, partir não do homem, mas da **necessidade social** que serve a sua profissão. Ortega y Gasset tratava daquilo que era demandante pela contemporaneidade para o ofício de bibliotecário, daquilo que era determinado de fora para dentro da profissão, e explicava: “esta necessidade, como tudo que é propriamente humano, não consiste em uma magnitude fixa, mas é, essencialmente, variável, migratória, evolutiva; em suma, histórica” (ORTEGA Y GASSET, 2006, p. 16).

A visão retrospectiva de Ortega y Gasset expõe a tensão existente entre a entropia do tempo histórico e a entropia da ação humana, e a natural disputa entre a ordem e a desordem. Oitenta anos depois da divulgação do seu texto, mais uma vez nos encontramos com o livro em uma perspectiva histórica. Passado o

3 Esta *quotation* atribuída a Charles Darwin é na verdade uma apreensão não literal de Megginson datada de 1963 ao pensamento de mudança apresentado por Darwin em *Origem das Espécies*: Cf. Leon C. Megginson na convenção da Southwestern Social Science Association. Veja a citação original: According to Darwin's Origin of Species, it is not the most intellectual of the species that survives; it is not the strongest that survives; but the species that survives is the one that is able best to adapt and adjust to the changing environment in which it finds itself. Disponível em: <<http://quoteinvestigator.com/2014/05/04/adapt/>>. Acesso em: 19 maio 2017.

turbilhão causado pela introdução da tecnologia digital e pela instalação social do ciberespaço, urge aos contemporâneos redefinir o que é o livro para nosso tempo.

Esta função viva, histórica, dinâmica nos desafia a abolir formatos pré-existentes, aceitar novas interfaces como forma de prover o acesso aos registros do conhecimento.

Livro é mutante, dinâmico e instável, mas mesmo assim insistimos em ancorar nosso entendimento à forma de sua última interface datada do século XV. É esta âncora que nos prende firme ao solo renascentista e que muitas vezes nos impede de navegar no mundo contemporâneo. Para uma reflexão produtiva, é necessário que nos desvistamos da imagem historicamente construída na época da introdução dos tipos móveis e da emergência da comunicação impressa, e acrescentemos à equação explicativa do livro as novas variáveis que o tempo presente impõe. Desse modo, para fins deste trabalho, quando nos referirmos ao conceito livro, não estaremos tratando do objeto tradicional, mas da interface multiplexa de registro do conhecimento, vislumbramos o voo da crisálida no desafio lançado por Ortega y Gasset aos seus contemporâneos:

Chega o momento de se haver com o livro na condição de conflito. É aqui, portanto, onde vejo surgir a nova missão do bibliotecário, incomparavelmente superior às anteriores. Até hoje ela tem se ocupado principalmente do livro como coisa, como objeto material. A partir de hoje terá que cuidar do livro como função viva (ORTEGA Y GASSET, 2006, p. 39).

A diferença mais eloquente entre os demais seres da criação e o homem está nesta função viva que nos faz produzir estoques de informação (BARRETO, 2000). Somos a única espécie que registra sua própria memória em interfaces documentais. Criamos próteses mentais ou exomemórias – manuscritos, impressos, objetos, arquivos digitais – com o fim deliberado de desonerar o aparelho cerebral, armazenar inteligência para fins científicos, econômicos, sociais e para geração de conhecimento novo.

O livro da era tecnológica perdeu-se do seu formato original. Não adormece mais nas estantes das bibliotecas. Está circulando em meio digital na web, nas redes sociais, nas organizações, nos objetos e nas coisas. O documento e a documentação também se desprenderam dos formatos analogicamente suportados e circulam nos mesmos caminhos do livro.

O documento e o livro são instrumentos evolutivos, complexos e instáveis, *"una extensión de la memoria y de la imaginación"* (BORGES, 1979, p. 165). Como defendeu Jorge Luiz Borges: As exomemórias ou interfaces da inteligência são extensões planejadas para projetar para a posteridade nossas realizações, para comunicar assincronamente de humanos para humanos, nossos avanços e falhas, e, para potencializar nossas virtudes.



As exomemórias são um tipo de instrumento do tempo, entidades artificiais e exteriores à natureza humana, criadas para potencializar as capacidades das inteligências para além dos limites impostos pela condição física e espiritual dos seres humanos, para ampliar nosso controle sobre o mundo natural e para o aproveitamento eficiente dos recursos oferecidos pela natureza.

Os instrumentos podem ser técnicos – quando destinados a potencializar as capacidades físicas e realizar tarefas impossíveis aos seres humanos em seu estado físico natural, como, caneta, automóvel – e os instrumentos lógicos – quando criados com o objetivo de potencializar as capacidades da inteligência humana. Atuam sistematizando e organizando tarefas impossíveis à execução dos indivíduos no campo lógico, e.g., a escrita, a ética, a matemática. Estas ferramentas trabalham em estreita associação com os instrumentos técnicos, diz-se desta associação instrumentos **tecnológico-lógicos**. Um documento é um bom exemplo de instrumento tecnológico composto por uma parte técnica (suporte) e uma outra lógica (escrita), destinado a potencializar a tarefa de preservação e perpetuação da memória e do conhecimento entre os indivíduos de uma sociedade local ou global.

As exomemórias já não são mais o livro, tampouco o documento analógico. Agora as exomemórias encarnam uma entidade, uma via de tradução entre a inteligência humana e o registro de seu conhecimento, aquilo que representa, que o interfaceia, o ser linear que se apresenta como zona de tradução, a função viva que Ortega y Gasset (2006) intuía e procurava definir na *Missão do Bibliotecário*, mutante, multiplexa que evolui da fala e do documento. É esta zona de tradução o objeto de interesse da curadoria.

Interface pode ser entendido como dispositivo desenvolvido com a finalidade de viabilizar a conexão entre inteligências, para operar a **comunicação** entre duas partes distintas e que não podem se conectar diretamente no tempo e espaço. Uma fronteira que define a forma de comunicação entre duas entidades. Para Lévy (1993, p. 181):

Interface é uma superfície de contato, de tradução, de articulação entre dois espaços, duas espécies, duas ordens de realidades diferentes: de um código para outro, do analógico para o digital, do mecânico para o humano [...]. Tudo aquilo que é tradução, transformação, passagem, é da ordem da interface.

Em um entendimento mais geral, interface é descrita por Johnson (2001, p. 17) como um espaço (zona):

A interface atua como uma espécie de tradutor, mediando entre as duas partes, tornando uma sensível a outra. Em outras palavras, a relação governada pela interface é uma relação semântica, caracterizada por significado e expressão.

Dispositivo com finalidade de conexão entre dois equipamentos que não possuem as mesmas funções. Ambiente de interação homem/máquina em qualquer sistema de informática ou automação. O projeto de uma interface eficiente, fácil de manusear e amigável (*user-friendly*) é um fator importante no planejamento de qualquer sistema de comunicação não-presencial. Ponto em que dois sistemas de características diferentes se interligam, sendo a informação a transmitir convertida de acordo com um conjunto de pressupostos e de regras conhecido de ambos. Também designa o conjunto de regras e pressupostos em si. São sinônimos de Interface conexão, área compartilhada, área de interação, campo de interação, dispositivo para troca de informação, ligação entre sistemas.

O trabalho de sistematizar a documentação realizado por Paul Otlet constituiu-se, em seu tempo, num esforço de releitura da missão histórica do livro, traduzindo para seus contemporâneos uma nova forma de perceber as interfaces do conhecimento que se multiplicavam durante a Revolução Industrial, diversa, multiplexa, enfim, hodierna, na qual o conceito aprisionado pela natureza crisálida do texto impresso encontrava sua metamorfose. De modo análogo, enxergamos o esforço dos gestores de dados contemporâneos que vêm desenvolvendo a curadoria de dados digitais como um aparelho social de gestão dos objetos da cultura, do patrimônio e da memória.

Um novo ciclo de inovação se aproxima, carreando transformações radicais que prometem ampliar o poder, a produtividade e a eficiência das organizações. Na primeira fase da Internet, apenas pessoas e organizações estavam conectadas, o mundo contemporâneo prepara-se para um novo salto quântico: a conexão das coisas, objetos domésticos, automóveis, serviços públicos, edifícios, entre outros. Em pouco tempo, tudo estará interligado, emitindo informações e gerando dados, e quanto mais tecnologizada estiver a organização, maior será sua capacidade competitiva. Neste momento, ser digital deixará de ser transitório e se tornará uma condição nativa.

Talvez no futuro os historiadores do conhecimento descrevam o tempo presente como o momento histórico da monetização em escala do conhecimento, quando ele se metamorfoseou tornando-se produtor de excedente e lucro, quando adquiriu as condições necessárias para funcionar como *commodities*, agregou valor e passou a ser gerido, estocado sem perda de qualidade e aproveitado como objeto de troca intensiva para fins de mercado.

Há alguns anos a humanidade se deu conta de que as organizações, através de suas plataformas de serviço, vinham, crescentemente, gerando dados sobre sua clientela. Percebemos que o conhecimento havia se tornado um novo tipo de ativo econômico, uma forma de recurso capaz de gerar novos recursos na forma de conhecimento. Mais: quando este conhecimento era utilizado de forma eficiente, era capaz de gerar economia inteligente. Justificados pela necessidade

e oportunidade de prestar atendimento personalizado aos clientes, com o auxílio da inteligência cognitiva as organizações criaram bancos de dados que qualificavam perfis de sua clientela.

A análise do padrão de crescimento do mercado de informação coloca as organizações diante de uma nova realidade que se aproxima rapidamente. O mundo se prepara para ampliar sua infraestrutura de processamento, tráfego e armazenamento de dados. Para se ter uma ideia do volume de investimentos nesta área, somente uma empresa de telefonia que atende a 97,2 milhões de clientes no Brasil, investiu, entre 2012 e 2017, quase 45 bilhões de Euros na ampliação de sua infraestrutura de rede<sup>4</sup>. Está claro também que esta inversão de capital ainda não é suficiente para atender as necessidades de gestão que o **Bigdata** vai demandar. É necessário o desenvolvimento de uma nova categoria de profissionais hiperespecializados, que em parte vem sendo suprida por profissionais da informática, da engenharia e de gestores de informação que por hora são denominados de curadores digitais.

*Paripassu* ao desenvolvimento de uma mentalidade renovada carreada pela Tecnologia da Informação, as formas tradicionais de gestão e uso do conhecimento permaneceram, por longo tempo, ativas e relevantes. Este fenômeno, sedimentado pelas gerações, obedece às regras naturais da vida em sociedade. O mundo analógico conviveu com o mundo digital por longo tempo em um processo que permitiu que aos poucos o modo digital se consolidasse, enquanto os processos analógicos gradativamente perdiam sua vigência, se retirando de cena com a geração anterior.

No mundo ocidental – desde a antiguidade até o final da Idade Média – a gestão do conhecimento era um privilégio, concedido a poucos, especialmente aquelas pessoas ligadas à igreja e/ou que dominavam o código escrito. Não obstante a abertura de pensamento promovida pelo escolasticismo até o final da Idade Média, somente aos iniciados, herdeiros de Levi<sup>5</sup> era permitido refletir, formar opinião e comunicar avanços. Fora desta circunscrição intelectual, qualquer produto de mentalidades independentes corria o risco de ser vista como heresia, a exemplo de Nicolai Copérnico, Galileu Galilei e Giordano Bruno, que tentavam consolidar a teoria do sistema heliocêntrico.

4 A perfil econômico Telefônica Brasil. Disponível em: <<http://telefonica.mediacom.com.br/pt/Empresa/Perfil.aspx>>. Acesso em: 18 maio 2017. Veja ainda entrevista com José Maria Alvares-Pallete Lopes Presidente do grupo Telefônica (LOPES, 2017).

5 Segundo a tradição judaica, levita era um membro da tribo de Levi com tarefas especiais, separada das outras tribos e mantinha em funcionamento os sacrifícios no templo de Jerusalém. Os levitas eram auxiliares dos sacerdotes. O Êxodo do Egito, foi liderado por Moisés e Arão que eram membros desta tribo, e no deserto eles eram incumbidos de zelar pelo tabernáculo, isto é desmanchar e montá-lo em outro lugar (Números 1:47-54). Os levitas não possuíam terra "porque o Senhor Deus de Israel é sua herança" (Deuteronômio 18:2) Moisés nomeou os descendentes de Levi, filhos de Aarão, como aqueles que teriam a permissão de realizar sacrifícios e adentrar o tabernáculo, e permanecer na presença à Arca da Aliança. Assim, na tradição cristã, por via da tradição hebraica, era natural que os sacerdotes se considerassem o trato do conhecimento como divino, e por conseguinte, reservado aqueles que Deus havia determinado.

Para uma compreensão mais clara deste fenômeno, podemos observar, com um necessário distanciamento histórico, o que ocorreu no período do Renascimento europeu, que teve lugar entre fins do século XIV e o fim do século XVI. O período foi marcado por radicais transformações na cultura, sociedade, economia, política e religião e pela transição do feudalismo para o capitalismo mercantil. Descrito em meados do século XIX por Burckhardt (2009) como o período de **descoberta do mundo e do homem**, o Renascimento é comumente relacionado à cultura e a forma como a humanidade se relacionava com o conhecimento bem como aos efeitos produzidos nas artes, na filosofia e nas ciências no mundo ocidental.

Originário da Toscana italiana, o Renascimento cultural se espalhou para o resto da península Itálica, e de lá avançou como uma febre sobre as demais nações europeias. Na Alemanha, o Renascimento permitiu o surgimento da imprensa e o cisma religioso que resultou no protestantismo, câmbio de largo impacto que ainda hoje repercute na vida das sociedades modernas.

Este processo deflagrou a emergência do que historicamente ficou convenido como **mundo moderno**, consolidando a ruptura com as estruturas medievais, e permitindo, em um segundo plano, o desenvolvimento do valor mais relevante do Renascimento: o pensamento humanista. Com o humanismo, o mundo do conhecimento abandonou progressivamente o método escolástico medieval<sup>6</sup>, criando as condições necessárias para o surgimento, no século seguinte, do pensamento analítico, apresentado ao mundo por René Descartes.

Um resultado clássico decorrente do Renascimento e da emergência do pensamento humanista é classificado pela história da ciência como Revolução Científica, expressão utilizada pela primeira vez em 1939 (KOYRÉ, 1939). Esta revolução foi marcada por um excepcional florescimento econômico e científico, pelo surgimento da imprensa, pela eclosão do hermetismo e da Reforma Protestante. Entre os séculos XVI e XVIII, a ciência, que até então se desenvolvia atrelada aos interesses da teologia, passou a valorizar a análise das evidências empíricas, o uso da razão individual e do senso crítico na construção do conhecimento estruturado experimental.

O humanismo ampliou o senso crítico e trouxe o homem para o centro dos interesses da ciência. Nesse contexto, o novo senso crítico permitiu aos homens da ciência observar mais atenta e rigorosamente os fenômenos naturais e, à ciência, mudar seus modos e função social. Reconfigurada, a ciência passou a

6 Escolástica (do latim *scholasticus*, instruído, aquele que pertence à escola) surgiu nos bancos das universidades medievais com o objetivo de diminuir as contradições que se manifestavam no confronto da fé cristã com o sistema de pensamento racional. Foi entre os séculos IX e XVI o método de pensamento crítico dominante no ensino nas universidades medievais europeias. Limitava-se basicamente à consulta às autoridades do passado, principalmente Aristóteles e os primeiros religiosos da Igreja, e ao debate das diferenças entre os autores e comentaristas. A partir da comparação entre o texto da obra e os documentos a ela relacionados, especialmente documentos da igreja e análises de estudiosos anteriores, se produzia as *sententiae*, curtas sentenças nas quais eram listadas as discordâncias entre fontes diversas, acerca dos temas tratados na obra em estudo (MACIEL, [200-]).

servir às novas demandas que surgiam, convergindo seus interesses com os da sociedade para um campo liberto das influências teológicas do mundo medieval.

Nesse período, Johannes Gutenberg introduziu o tipo móvel, e a imprensa se disseminou, cumprindo fundamental papel na difusão do conhecimento. Os avanços da revolução científica foram incentivados pelos reformistas, que viam na ciência uma forma de reificar a existência de Deus e propagar a reforma. O hermetismo, a seu turno, coroou a Revolução Científica, defendendo um corpo de postulados analíticos que exaltavam a concepção quantitativa do universo. A popularização da linguagem matemática foi decisiva para o desenvolvimento de uma mentalidade metodológica que avançava progressivamente em direção do rigor e da crítica. Essa circunstância cambiou definitivamente o modo de fazer ciência na Europa, com poderoso impacto nos avanços que se seguiram nos séculos posteriores.

A Revolução Científica não se deu por geração espontânea, antes disso, foi um movimento articulado que eclodiu em diversos pontos da Europa com uma motivação universalizada. É comum a visão estanque que observa a emergência do renascimento e do humanismo, o desenvolvimento tecnológico que levou às grandes navegações, ao surgimento da imprensa, como eventos históricos separados, com pouca relação entre si. Ao contrário disso, a Revolução Científica é produto de uma nova ordem econômica e social, produzida pela ressonância mórfica de uma nova noosfera<sup>7</sup>, resultado da inquietação de uma sociedade em metamorfose, refundada por princípios humanísticos, e fertilizada pela oferta de novas ideias. A exemplo do que se deu com o pensamento humanista que, a seu tempo, inspirou e mobilizou as mentes do Renascimento, o Positivismo atuou no mundo contemporâneo, semeando novas ideias e estabelecendo uma cultura articulada de largo domínio e impacto social.

## O IMPÉRIO DO POSITIVISMO

A doutrina filosófica, sociológica e política positivista surgiu na França como resultado da crise social e moral no começo do século XIX e da emergência da sociedade industrial. Segundo Zuin (2007), os processos de transformação socioeconômica, as crises e as rupturas da ordem política revelavam um mundo que se desfazia, possibilitando o “[...] surgimento de novos debates sobre a missão dos

<sup>7</sup> Sheldrake (2014) esclarece que, Ressonância Morfica é o processo responsável por essa coletivização da informação, por meio dela, as informações se propagam no interior do campo mórfico, alimentando uma espécie de memória coletiva. C.F. e Gnoosfera ou Noosfera tem sua raiz no termo “gnosio” (conhecimento Global). Segundo Marc Halévy, citando Teilhard de Chardin, é essa “camada de Saberes e de conhecimentos que cobre toda a terra e suas redes, e se sobrepõe a sociosfera.” A Noosfera, ou esfera pensante superposta coextensivamente [...] à biosfera, designa genericamente o conjunto das redes de ideias e de conhecimentos em que se desenvolvem os processos de criação, memorização, transformação e transmissão de noemas. É o lugar de suas proliferações autônomas. É uma camada imaterial plantada acima da sociosfera humana, mas distinta dela (HALEVY, 2010, p. 14).

intelectuais na nova ordem social e política” (ZUIN, 2007, p. 67). O positivismo se integrava em um desses contextos históricos, e propôs uma nova classificação do conhecimento associando uma interpretação das ciências a uma ética humana radical. Para os positivistas, o progresso da humanidade dependia exclusivamente dos avanços científicos; assim, eles desconsideravam qualquer forma de conhecimento que não pudesse ser comprovado cientificamente.

Os bibliotecários foram historicamente serventuários do conhecimento, e por extensão, associados da igreja. No ambiente intelectual renovado pelo positivismo, todavia, profissionais progressistas passaram a se autodenominar Servos dos Servos da Ciência (*Scientia Servus Servorum*), uma ironia alusiva à expressão latina *Servus Servorum Dei* pela qual, no final do século XIX, os Papas se identificavam nas bulas pontifícias. O trocadilho escolhido pelos bibliotecários guardava um duplo sentido: o primeiro marcava o território positivista, e o segundo, a escolha laica dos bibliotecários que, àquela época, passavam a manter estreita relação com os cientistas, engajados na construção da Sociedade Industrial<sup>8</sup>. Um bom exemplo da prática positivista de se louvar a condição serviçal dos bibliotecários brasileiros pode ser apreciado na nota biográfica de Taunay (1942), composta para introdução da edição portuguesa da *História natural do Brasil* de Jorge Marcgrave. Nesse ensaio, Taunay (1942, p. 3) explica que somente conseguiu acessar a rara bibliografia “*mercê do extremo serviçalismo*” do bibliotecário e do superintendente da sala de leitura da Biblioteca do Congresso em Washington. O qualificativo **serviçalismo**, hoje pejorativo, à época soava elogioso para os bibliotecários.

A postura dos progressistas comtianos aclimatou-se nas latitudes tropicais do Brasil, assumindo posição oposta àquela que era utilizada originalmente pelos bibliotecários europeus. Até bem pouco tempo, alguns bibliotecários assumiam voluntariamente essa condição de serviçalismo intelectual que resumia um espírito de corpo, um modo de pensar de uma época. Este *modus* colocava os bibliotecários em posição de subalternidade frente aos **verdadeiros** servos da ciência: os cientistas, limitando seu público, desta forma, a uma estreita faixa da elite culta. Essa leitura, em um só ato, assumia a condição tecnicista, transferia aos pesquisadores a obrigação da reflexão e da crítica e abrigava-os em uma zona de conforto que resumia sua função social à custódia e ao controle bibliográfico.

8 Sobre este tema ver: Castro (2000). Ver também Castro e Ribeiro (2004). “Deste modo, são exigidos novos perfis ao profissional, aqueles que dessem conta de controlar a produção bibliográfica nacional, em especial, a técnico-científica. Daí, os bibliotecários passaram a adaptar a insígnia ‘Servos dos Servos da Ciência’: Os bibliotecários entendiam a expressão como algo positivo, na medida em que se consideravam como auxiliares dos cientistas, a quem estes recorriam quando necessitavam de informação para o desenvolvimento de seus estudos e pesquisas. Enfim, o ato de servir à ciência era um sinal de modernidade bibliotecária, ou seja, uma atividade que requeria especialização em um dado campo. Desse modo, essa expressão marcava a diferença entre o bibliotecário generalista do qualificado, especializado. Mas o que significa servo? Sujeito, vassalo, dependente, subalterno, inferior, pessoa a serviço de alguém, o que podemos concluir?” (CASTRO; RIBEIRO, 2004, p. 44).



Esta tendência, magnificada com o evento da especialização, acabou afastando os bibliotecários da função social original que historicamente havia conduzido os gestores do conhecimento rumo ao desenvolvimento de instrumentos técnicos e lógicos aviadores do processo de *sinapse* que, por sua vez, promoviam a conexão inteligente e eficiente entre os registros do conhecimento e os usuários. No vácuo reflexivo, novos operadores do conhecimento passaram a desenvolver ferramental para gestão de ativos e passivos do saber, utilizando-se para este fim a nova ciência da informática auxiliada pela tecnologia da informação.

No final do século XIX, o mundo ocidental experimentava os efeitos do *clímax* produzido pela Era Industrial. A economia exuberante impactava sensivelmente nos mais diferentes aspectos da sociedade moderna. O capital que migrava da produção manufatureira formava impérios locais, imprimia uma nova forma de vida, tanto nos centros urbanos quanto no mundo rural. Nesse período o mundo também se preparava para duas grandes guerras que se sucederam entre 1914 e 1945.

A epifania econômica e o surgimento da sociedade de consumo levaram a um dos mais profícuos períodos de desenvolvimento tecnológico da humanidade. Nesse período surgiu a eletricidade, os veículos terrestres e aéreos motorizados, o rádio e o telégrafo, o trem e a imprensa moderna. Inspirados pelo apogeu econômico da Europa no século XIX, coube, a certo modo, à literatura do final do século XIX, delinear a imagem do futuro que se aproximava por meio da ficção. Diante de um cardápio tecnológico tão variado, os ficcionistas como H. G. Wells, Pierre Giffard, Albert Robida, Júlio Verne e Edgar Allan Poe, deram à literatura a oportunidade de usar a criatividade para antever o inusitado que se descortinava por trás da linha do conhecimento.

A conjugação entre a expansão do modo industrial e o ciclo dos conflitos universais exigiu do mundo moderno a produção de conhecimento em uma escala ainda desconhecida. No *Post bellum* um *corpus* de novas conquistas tecnológicas e um admirável estoque de dados científicos produzidos durante as duas grandes guerras, fertilizaram mentes inovadoras nos mais variados campos do conhecimento.

A gestão do conhecimento registrado emergia como um problema de amplo interesse social, atraindo a atenção de *decision-makers*. Depois das descobertas do matemático Alan Turing, os estrategistas perceberam o poder que se agregava ao processamento e compreensão algorítmica. A visão complexa de Turing apresentou ao mundo uma nova forma de leitura da realidade que vinha dos dados, e que ultrapassava largamente aquela outra da visão puramente analítica. Mais que uma leitura sintática da informação, os algoritmos liam os dados e apresentavam tendências e estatísticas, e essas tendências construíam a realidade semântica.

A sistematização do conhecimento amalhado exigia o desenvolvimento de métodos organizacionais de alta performance. Nesse período se destacam os esforços de Paul Otlet (Europa) e Vanevar Bush (EUA), ambos de granular relevância para a construção da Ciência da Informação.

Nas obras *Traité de documentation* (OTLET, 1934) e *Monde: essai d'universalisme* (OTLET, 1935), descrevia não apenas uma nova forma de recolher e organizar o mundo do conhecimento, mas apontava o rumo futuro que a documentação deveria percorrer. Uma década depois, Bush (1945) publicou o artigo *As we may think*, apontando os problemas decorrentes do volume e do valor da informação mantida secreta até a segunda Guerra Mundial e que, então, seria colocada à disposição do mundo.

Em seu célebre texto, Bush (1945) retoma a ideia da formação de uma biblioteca universal, já proposta de outro modo, décadas antes, por Paul Otlet (1934). O desafio que se apresentava era o de gerenciar de forma inteligente grandes massas de conteúdo, padronizando sua representação, de modo que a recuperação de dados fosse mais eficiente e produtiva.

Bush estava preocupado em criar um mecanismo que armazenasse e facilitasse a pesquisa de toda essa produção tecnocientífica que crescia em ritmo acelerado.

Nesse artigo – inspirado pelo mesmo espírito que iluminava a documentação de Otlet – Bush (1945) descreve o Memex (memory+index), dispositivo a partir do qual os profissionais poderiam consultar seus pares e buscar pareceres e processos semelhantes aos seus. Nesse prodigioso invento que nunca chegou a ser construído, as pesquisas se dariam pela semelhança e associação dos termos buscados. Além da pesquisa associativa, que é muito semelhante ao modo como o pensamento humano funciona, o conteúdo estaria disposto pela similaridade por meio de elos que ligariam os assuntos uns aos outros. Assim nasceu também o conceito de hiperlink, que atualmente é imprescindível para se navegar pela internet.

Além dessas propostas visionárias, a máquina ainda possibilitaria a edição e modificação do conteúdo pelo usuário. Em vez de serem guardadas em bancos de dados, como as da internet, as informações do Memex estariam gravadas em microfilmes e fotografias que seriam manipulados diretamente pelo usuário.

No *Mundaneum*, Paul Otlet (1929) tentava, a seu turno, reunir referenciais do universo do conhecimento produzido até sua época. Neste esforço, amalhou uma quantidade insondável de referencias, que, em que pese não oferecerem acesso aos dados referenciados, permitiam que os cientistas tivessem acesso ao universo representado, e então poderiam buscar as fontes correspondentes.

O acesso efetivo, em larga escala, aos registros do conhecimento somente foi possível com o advento da reprografia e do microfilme. Todavia, o volume e a qualidade deste acesso somente seria otimizado depois da introdução dos meios eletrônicos surgidos durante a segunda grande guerra.

Evidências históricas mostram que as nações que participaram ativamente em conflitos bélicos universais experimentaram uma combinação de desenvolvimento industrial, acúmulo de capital econômico, formação de centros de pesquisa e universidades avançadas, associada à consolidação de uma elite científica. Este era o caso da América de Busch. O alinhamento dessas condições criou um ambiente privilegiado de produção e gestão do conhecimento que resultou no acúmulo de valioso capital intelectual. Este processo histórico pode ser observado, tanto no século XV, influenciando o Renascimento, quanto no final do século XIX, no apogeu da Revolução Industrial que antecedeu as duas grandes guerras mundiais.

Nações que não partilharam desta experiência produziram experiências sociais pobres no tocante à produção e gestão do conhecimento. Excluído dos grandes eventos políticos, do cenário econômico mundial, bem como dos grandes conflitos hegemônicos globais, o Brasil acabou se separando do bloco universal que expandia as fronteiras do conhecimento, e por esta razão também ficou fora do grande salto da ciência e da documentação. Isolado, somente em 1915 surgiu o primeiro curso de Biblioteconomia no Brasil, representando uma tímida participação neste grande fenômeno de deriva que acabou gerando em nós o efeito Madagascar<sup>9</sup>.

O **efeito Madagascar** pode ser definido como uma forma de endemismo intelectual que acabou propiciado o isolamento dos profissionais que tinham por ofício a gestão do conhecimento no Brasil. A ideia de endemismo intelectual é aplicada a grupos que se desenvolveram de modo endógeno numa região restrita. Em geral, resulta do isolamento que propicia a origem de tendências regionais que evoluem com formas próprias desconectadas do sistema global. O endemismo é causado principalmente por mecanismos de isolamento que impedem ou dificultam as trocas culturais. Um dos resultados do endemismo é a autossuficiência e o pseudoentendimento de eficiência.

<sup>9</sup> É imperativo notar que o país havia permanecido em isolamento cultural por quase todo o período colonial, as políticas lusitanas não permitiram o desenvolvimento da educação nas colônias, temendo que a instrução pudesse formar mentes independentes e as levantar contra a metrópole. Como parte desta política livros e bibliotecas foram vistos como potenciais focos de subversão, uma vez que nestes lugares, a leitura, a cultura e o conhecimento abriam o entendimento e propiciavam a reflexão sobre a condição humana. Esta política não era uma característica do processo colonial brasileiro, na América do Norte o combate à instrução dos colonos também foi vigorosa. Hallewel (2005) apresenta-nos em sua História do Livro no Brasil uma citação a Sir William Berkley, Governador da província da Virgínia, em 1671, que em uma correspondência oficial explícita de modo claro este pensamento. Em sua fala Berkley dava graças a Deus por não ter havido na América até sua época "*escolas livres nem imprensa, e espero que não a tenhamos por trezentos anos ainda.*" concluía enfático o velho governador: "*Porque aprender trouxe ao mundo a desobediência e a heresia*". Esta cultura colonial talvez explique um pouco o atraso da educação brasileira que efetivamente foi liberada após a Revolta do Porto de 1820.

Muitos bibliotecários e arquivistas, presos radicalmente ao corporativismo e ao ofício tecnicista, negaram a inovação e não perceberam o momento histórico que se encenava diante dos seus olhos. Um ramo dos vanguardistas que se associaram aos matemáticos, físicos, engenheiros e outros cientistas engajados no esforço de guerra, acabou construindo sobre a base firme da documentação, a moderna Ciência da Informação, outro ramo destes pioneiros, servindo-se em larga medida de instrumentos de organização do conhecimento como tesouros, linguagens controladas, métodos de classificação do conhecimento, entre outros – desenvolvidos por bibliotecários – criaram as bases da informática e da ciência da computação moderna.

Nos países desenvolvidos, o avanço tecnológico acabou reificando as responsabilidades do Estado para com os bens do patrimônio e a memória. Como reflexo desta consciência pública, as instituições de memória passaram a receber mais atenção e recursos. No final do milênio, a mais significativa parte do mundo desenvolvido já contava com políticas públicas que abraçavam, indistintamente, a memória tradicional e a digital, como depositárias dos bens da cultura universal. A exemplo do que ocorreu em seu tempo com a documentação, deste esforço inicial resultou o desenvolvimento dos padrões e protocolos que ainda hoje regem o funcionamento dos instrumentos tecnológicos que viabilizam a troca de dados no mundo inteiro.

Para entender as circunstâncias que propiciaram os avanços tecnológicos atuais, e a evolução dos instrumentos de recuperação, processamento, distribuição e acesso a dados em ambientes de alta performance das redes, é necessário entendê-las como parte de um intrincado processo de complexificação social decorrente do avanço econômico, acúmulo de capital e oferta de recursos. Lançar luzes históricas sobre este processo permite-nos vislumbrar os avanços experimentados hoje como resultado do esforço intelectual semeado por nossos antepassados em diferentes momentos do nosso passado histórico.

## **SOBRE ESTE NOVO UNIVERSO E AS NOVAS FORMAS DE ORDENAÇÃO, RECUPERAÇÃO E ACESSO AO CONHECIMENTO**

O novo limiar tecnológico que se anuncia se apropria de nossa agenda, os hodiernos já veem um novo teatro se apresentar, no senso do lugar público de onde se vê, é o *locus* condensado da vivência das ambiguidades e paradoxos, um novo limiar tecnológico que certamente exigirá novas e mais complexas ferramentas para ordenação do espaço da inteligência compartilhada. A curadoria

digital vem emergindo como uma dessas ferramentas, e certamente cumprirá em nossos dias o papel que a documentação ocupou no século XX.

Em meados do século XX, o advento do desenvolvimento da computação e, mais tarde, na virada do século, a integração dos computadores em redes de alcance universal, apresentaram ao mundo uma nova escala de gestão e uso de informação que confirmaram as mais ousadas predições de Otlet e de Bush.

Com o surgimento em 1994 da *World Wide Web Consortium*, proposta por Tim Berners-Lee, um enorme volume de informação registrada em meio digital passou a trafegar nos ambientes distribuídos da web.

Os ambientes modificados pela superexposição aos novos instrumentos de criação geram demandas de conhecimento cada vez mais complexas. Esta circunstância exigiu, mais uma vez, o esforço conjugado de cientistas para desenvolver ferramentas capazes de ordenar a magnífica noosfera que emergia do encontro de inteligências interconectadas em rede.

Assim, a tecnologia da informação ofereceu o diferencial, o ponto fora da linha, aquilo que fugia ao acervo da experiência social, o que escapava ao horizonte das expectativas. No princípio da revolução da tecnologia digital, havia um misto de sentimentos que alternavam entre euforia e medo. Enquanto alguns se apressavam em se postar como mensageiros do novo tempo, outros se entrincheiravam contra a tecnologia que era vista e sentida como um corpo estranho no mundo de alma analógica.

A evolução e a mudança, contudo, como tudo que é natural, vinham do distinto, do incomum, do inesperado. Da tensão resultante das forças que competiam, emergia o vetor essencial que conduziu e manteve os câmbios em um campo sustentável e promissor. Assim, quando as forças inovadoras se estabeleciam, passavam a fazer parte do corpo formal da sociedade.

O diferente é, portanto, a força propulsiva da mudança que experimentamos enquanto humanidade. Somos todos, por natureza, gênero *Homo*, espécie *sapiens*. A humanidade, entretanto, não se conquista por nascimento, é um estado de espírito que se alcança socialmente com esforço e determinação.

A tecnologia, a expressão mais pretérita da racionalidade, foi criando instrumentos técnicos e lógicos com que nos fizemos *homo* e *sapiens*. A assunção desta assertiva implica em aceitar que nada expressa mais a condição humana que a tecnologia, e se a tecnologia nos faz humanos, o digital pode ser considerado o estado da arte da humanidade.

Seria a curadoria digital herdeira, em linhagem direta, da documentação otletiana, e seu **imago** no mundo contemporâneo?

## REFERÊNCIAS

ARISTÓTELES. **Política**. Brasília: UNB, 1997.

BARRETO, A. A. Os agregados de informação: memórias, esquecimento e estoques de informação. **DataGramaZero**, v. 1, n. 3, jun. 2000. Disponível em: <<http://www.brapci.ufpr.br/brapci/v/a/789>>. Acesso em: 27 maio 2017.

BÍBLIA, A. T. Deuteronômio 18:2. In: BÍBLIA. Português. **Bíblia Sagrada**. São Paulo: Ave Maria, 2010.

BÍBLIA, A. T. Números 1: 47-54. In: BÍBLIA. Português. **Bíblia Sagrada**. São Paulo: Ave Maria, 2010.

BORGES, J. L. El libro. In: BORGES, J. L. **Borges, Oral**. Buenos Aires: Emecé Editores: Editorial Belgrano, 1979.

BURCKHARDT, J. **A cultura do renascimento na Itália**: um ensaio. São Paulo: Companhia das Letras, 2009.

BUSH, V. As we may think. **Atlantic Monthly**, v. 176, n. 1, p. 101-108, 1945. Disponível em: <[http://worrydream.com/refs/Bush%20-%20As%20We%20May%20Think%20\(Life%20Magazine%209-10-1945\).pdf](http://worrydream.com/refs/Bush%20-%20As%20We%20May%20Think%20(Life%20Magazine%209-10-1945).pdf)>. Acesso em: 29 maio 2017.

CASTRO, C. A. **História da biblioteconomia brasileira**. Brasília: Thesaurus Editora, 2000.

CASTRO, C. A.; RIBEIRO, M. S. P. As contradições da sociedade da informação e a formação do bibliotecário. **RDBC: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 1, n. 2, p. 41-52, dez. 2004. Disponível em: <<https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/2079>>. Acesso em: 03 jul. 2017.

HALÉVY, M. **A era do conhecimento**. São Paulo: Unesp, 2010.

HALLEWELL, L. **O livro no Brasil**: sua história. São Paulo: Edusp, 2005.

JOHNSON, S. **Cultura da interface**: como o computador transforma nossa maneira de criar e comunicar. Rio de Janeiro: J. Zahar, 2001.

KOYRÉ, A. **Études Galiléennes**. Paris: Hermann, 1939.

LÉVY, P. **As tecnologias da inteligência**: o futuro do pensamento na era da informática. Rio de Janeiro: Editora 34, 1993.

LOPES, J. M. A. **VEJA**, Páginas amarelas, 10. ed., mar. 2017. (Entrevista com o presidente do grupo Telefônica).

MACIEL, W. **Escolástica**, [200-]. Disponível em: <<http://www.infoescola.com/filosofia/escolastica/>>. Acesso em: 29 maio 2017.

MEGGINSON, L. C. Lessons from europe for american business. **Southwestern Social Science Quarterly**, v. 44, n. 1, june 1963.

MOURÃO, G. **Ensaio sobre a solidão**. Rio de Janeiro: [s.n.], 2013.

ORTEGA Y GASSET, J. **Missão do bibliotecário**. Brasília, DF: Briquet de Lemos, 2006.

OTLET, P. **Cité Mondiale**. Geneva: World Civic Center: Mundaneum Union of International Associations, 1929.

OTLET, P. **Monde**: essai d'universalisme: connaissance du monde, sentiment du monde, action organisée et plan du monde. Brussels: Editiones Mundaneum, 1935.

OTLET, P. **Traite de documentation**. Brussels: Editiones Mundaneum, 1934.

SHELDRAKE, R. **Uma nova ciência da vida**. São Paulo: Cultrix, 2014.

TAUNAY, A. de E. Esboço biográfico de Jorge Marcgrave. In: MARCGRAVE, J. **História natural do Brasil**. São Paulo: Museu Paulista; Imprensa Oficial do Estado, 1942. p. I-IV.

ZUIN, J. C. S. A crise da modernidade no início do século XX. **Estudos de Sociologia**, v. 6, n. 11, 2001. Disponível em: <<http://seer.fclar.unesp.br/estudos/article/viewFile/412/1210>>. Acesso em: 17 abr. 2017.





# GESTÃO INTEGRADA DA INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA E O ACESSO ABERTO: ONDE ESTAMOS E ONDE PODEMOS CHEGAR

Fernando César Lima Leite  
Michelli Pereira da Costa





## INTRODUÇÃO

A capacidade de criação, compartilhamento e uso do conhecimento científico, e, conseqüentemente, a produção e uso de informação científica, é um dos fatores determinantes do desenvolvimento de uma nação. Durante a histórica *International Conference on Scientific Information*, no ano de 1959, Alexander Ivanovich Mikhailov, um dos principais estudiosos da informação científica, afirmou que o rápido progresso da ciência e tecnologia somente é possível com serviços de informação bem organizados (MIKHAILOV, 1959). De acordo com o autor, devido ao nível de desenvolvimento da ciência, à época, é evidente que muitas questões científicas e técnicas poderiam ser resolvidas com maior velocidade se dispusessem de acervos de informação amplos e sistemáticos sobre todos os campos do saber. Não resta dúvida de que os anseios demonstrados por Mikhailov ainda se fazem presentes neste início do século XXI.

São constantes as modificações e transformações sofridas pelo sistema de comunicação científica. Decorrente de fatores sociais, econômicos, legais e, em grande, parte tecnológicos, os processos e meios de promoção de fluxos de informação científica, quer do ponto de vista formal, quer do ponto de vista informal, têm sido forte e positivamente impactados pelo movimento de acesso aberto. Seus desenvolvimentos e avanços são percebidos em todo o globo, mesmo que sob diferentes gradações em razão de disciplinas ou região. Do ponto de vista de sua finalidade, acesso aberto tem a ver com esforços para promover e garantir fluxos de informação desimpedidos de modo que a produção do conhecimento científico possa ser facilitada e acelerada. Do ponto de vista de sua operacionalização, o acesso aberto lança mão de processos de gestão da informação científica, de modo que sua finalidade seja alcançada.

Este capítulo possui três objetivos. O primeiro é apresentar o modelo de acesso aberto que predominou no Brasil desde o ano de 2003, os caminhos percorridos e suas principais estratégias. O segundo objetivo é fazer uma leitura do acesso aberto sob a lente da gestão da informação científica, evidenciando a vinculação entre as duas práticas. Por fim, o terceiro objetivo é propor o caminho a ser seguido, alinhando soluções para problemas antigos com tendências.

## **SOBRE ACESSO ABERTO À INFORMAÇÃO CIENTÍFICA**

O uso das tecnologias no contexto da comunicação na ciência tem sido responsável por inúmeras transformações em seus processos. Lagoze e Van de Sompel (2001) afirmam que a introdução em grande escala das tecnologias no processo de comunicação científica gerou demanda do uso da web para a disseminação dos resultados de pesquisas. No bojo dessas mudanças, a utilização de recursos digitais beneficia o processo de comunicação, sobretudo, à medida que são aperfeiçoados e tornados mais ágeis os fluxos de informação e conhecimento científico e, além disso, quando se constata transformações essenciais em processos, elementos ou funções exercidas.

Nesse sentido, apregoa-se que um dos grandes avanços tecnológicos na comunicação científica foi a criação dos periódicos científicos eletrônicos. Embora sejam referenciados como inovadores no limiar da década de 90 do século passado, a estruturação e a lógica do modelo de comunicação tradicional, especialmente do sistema de publicações científicas até então predominante, sofreram poucas modificações significativas, o que significa que não houve uma transformação, mas sim uma modernização de processos de publicação científica. Dentre os principais benefícios desta modernização está, principalmente, a ampliação da possibilidade de acesso aos conteúdos.

Portanto, mesmo com o surgimento dos periódicos científicos eletrônicos, a hegemonia de editores científicos – que na realidade conduzem, com o aval da própria comunidade científica, o cerne do sistema de publicações científicas – e a distribuição de informação científica começam a ser alvo das transformações apenas recentemente. A própria comunidade científica passa a questionar a lógica do sistema de publicação científica tradicional, em que editores científicos comerciais retêm os direitos autorais patrimoniais, atribuem preços excessivos e impõem barreiras de permissão sobre publicações de resultados de pesquisas financiadas com recursos públicos, como discutem Odlyzko (2006) e Willinsky (2006), limitando a visibilidade e a circulação do conhecimento científico, conforme ilustrado na Figura 1.

Em última análise, dessa lógica decorrem dois problemas principais relacionados. O primeiro é a limitação dos impactos dos resultados de pesquisa, já que a limitação do acesso diminui consideravelmente as possibilidades de uso por outros pesquisadores, como será discutido a seguir. O segundo problema reside no fato de que, a despeito da modernização de processos de comunicação científica, pesquisadores encontram dificuldades para ter acesso a toda a informação que necessitam para desenvolver suas atividades.

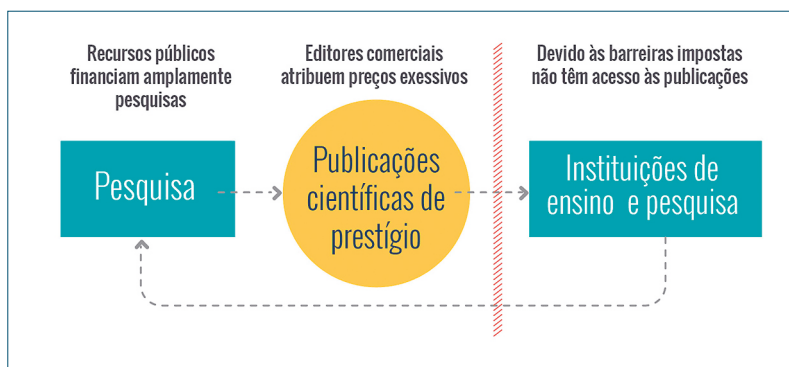


Figura 1 – Lógica do sistema de publicação científica tradicional  
Fonte: Leite (2011).

Constituído a partir dessa lógica, pode-se afirmar que o sistema de comunicação científica tradicional limita, mais do que expande, a disponibilidade e legibilidade da maior parte da pesquisa científica, conforme sugere Johnson (2002). A Figura 2, adaptada de Brody e Harnad (2004), ilustra o impacto limitado pelo acesso restrito a resultados de pesquisa, em que o acesso à informação ocorre apenas no momento de sua publicação, mediante pagamento.

Da insatisfação da comunidade científica com a lógica predominante do sistema aliada aos avanços tecnológicos, especialmente da tecnologia de arquivos abertos, culminou um esforço internacional de promoção da ampla circulação de resultados de pesquisa nas comunidades científicas, o movimento mundial em favor do acesso aberto à informação científica (BRODY et al., 2004; COSTA, 2006; BERLIN, 2003; JACOBS, 2006; WILLINSKY, 2006). Suber (2007) mantém uma linha do tempo dos principais eventos relacionados com o surgimento e avanços do acesso aberto no mundo. Para esse autor, o acesso aberto é sobre a literatura que é digital, online, livre de custos e livre da maior parte das restrições de direitos de cópia e licenciamento. O autor considera que o conceito central do acesso aberto está relacionado à remoção das barreiras de preço (custos de assinatura) e barreiras de permissão (restrições de direitos de cópia e licenciamento), de modo a tornar a literatura científica livre, tornando-a disponível com o mínimo de restrições.



Figura 2 – Acesso restrito: impacto da pesquisa limitado

Fonte: Adaptado de Brody e Harnad (2004).

De acordo com uma de suas principais definições, alcançada durante a realização da Budapest Open Access Initiative (2001), o acesso aberto é promovido por pesquisadores sem qualquer expectativa de pagamento. Nesse contexto, requer que a literatura científica esteja disponível livre e publicamente na Internet, de forma a permitir a qualquer usuário a leitura, download, cópia, distribuição, impressão, busca ou criação de links para os textos completos dos artigos, bem como capturá-los para indexação ou utilizá-los para qualquer outro propósito legal. Para tanto, é operacionalizado a partir de duas estratégias principais: auto-arquivamento em repositórios digitais e publicação de periódicos científicos de acesso aberto. O pressuposto de apoio ao acesso aberto requer que não haja barreiras financeiras, legais ou técnicas, além daquelas próprias do acesso à Internet. A única restrição à reprodução e distribuição e a única função do copyright neste contexto devem ser o controle dos autores sobre a integridade de sua obra e o direito de serem adequadamente reconhecidos e citados (BUDAPEST OPEN AC-

CESS INITIATIVE, 2001). Ao examinar as definições da declaração, Bailey (2006) chama a atenção para seus aspectos centrais, elencados a seguir:

- a) conteúdos em acesso aberto significa que estão livremente disponíveis;
- b) online significa que se trata de documentos digitais disponíveis na Internet;
- c) dizem respeito a conteúdos de natureza acadêmica/científica, ou seja, documentos administrativos, romances, revistas, livros de autoajuda e outros estão excluídos do conceito;
- d) os autores desses trabalhos não são pagos por seus esforços;
- e) uma vez que a maioria dos autores de artigos de periódicos científicos não são pagos e que tais trabalhos são acadêmicos/científicos, esses artigos constituem o principal tipo de material de acesso aberto;
- f) além da necessidade de atribuição de autoria e da manutenção da integridade da obra, os usuários podem copiar e distribuir os conteúdos sem qualquer restrição;
- g) há duas principais estratégias de acesso aberto: autoarquivamento em repositórios digitais e publicação de periódicos científicos de acesso aberto.

Segundo a mesma iniciativa, uma velha tradição e uma nova tecnologia convergiram para tornar possível o aparecimento de um bem público sem precedentes, que é a circulação livre e gratuita da literatura científica em escala global, Figura 3.

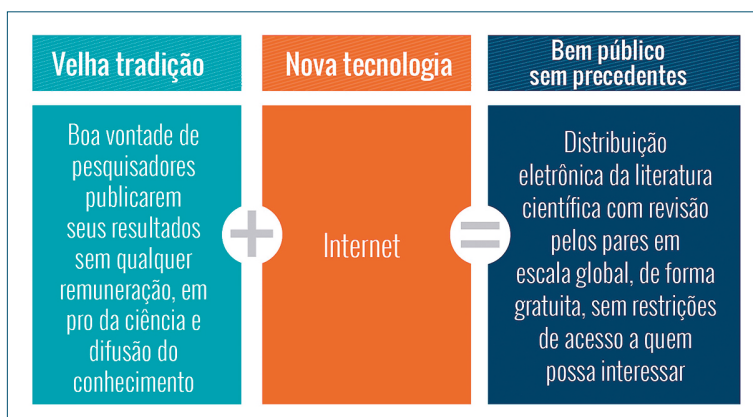


Figura 3 – Elementos do acesso aberto

Fonte: Adaptado de Budapest Open Access Initiative (2001).



O conceito de acesso aberto pode ser aplicado de duas maneiras distintas, conforme explica Morrison (2009). A primeira refere-se a trabalhos de acesso aberto. Os conteúdos são de acesso aberto quando forem, no mínimo, livres para qualquer um, em qualquer lugar para ler online (acesso aberto grátis), ou melhor ainda, livre para ler e livre para qualquer um reutilizar (acesso aberto livre). Morrison chama a atenção para o fato de que uma obra que tem seu acesso fechado determinado por direito de cópia pode se tornar de acesso aberto quando entrar em domínio público. O segundo caso de aplicação do conceito de acesso aberto refere-se ao processo, tal como a publicação. O termo acesso aberto, nesse caso, deve ser reservado para publicações de acesso aberto pleno, onde um determinado trabalho é feito em acesso aberto (grátis ou livre) como parte de um processo de publicação, sem atraso. Ambas as aplicações do conceito denotam oportunidades e benefícios, como discutidos a seguir.

O acesso aberto a resultados de pesquisa tem sido visto como fator que maximiza o acesso à pesquisa propriamente dita. Dessa maneira, aumenta e acelera o impacto das pesquisas e, conseqüentemente, sua produtividade, progresso e recompensas, conforme explicam Brody e Harnad (2004). Lawrence (2001) analisou 119.924 trabalhos apresentados em conferências na área de informática e demonstrou que a média de citações feitas a artigos *offline* era de 2,74. Em contrapartida, a média de citações a artigos disponíveis publicamente na rede era de 7,03, correspondente a um aumento de 336%. Como observa Lawrence, para maximizar o impacto, minimizar a redundância e acelerar o progresso científico, autores e editores deveriam tornar mais fácil o acesso aos resultados de pesquisa. Do mesmo modo, Brody et al. (2004) demonstraram que artigos disponíveis livremente recebem entre 2,5 e 5,8 mais citações do que artigos *offline*. A literatura registra resultados de diversos outros estudos, em diferentes áreas do conhecimento, que demonstraram em sua maioria o aumento das taxas de citação. Swan (2010), analisando 31 desses estudos, concluiu que, desses, 27 relataram vantagens positivas de citação e 4 demonstraram não haver vantagem ou então desvantagem de citação. A Tabela 1 ilustra o percentual de aumento das taxas de citação relatado pelos diferentes estudos analisados pela autora.



Tabela 1 – Tamanho da vantagem de citação de acesso

DISCIPLINA	% de aumento das citações com o acesso aberto
Física/ Astronomia	170 a 580
Matemática	35 a 91
Biologia	-5 a 36
Engenharia Elétrica	51
Ciência da Computação	157
Ciência Política	86
Filosofia	45
Medicina	300 a 450
Estudos da Comunicação (TI)	200
Ciências Agrícolas	200 a 600

Fonte: Adaptado de Swan (2010).

Um dos meios mais eficazes de facilitar o acesso à pesquisa é torná-la disponível livremente. Com isso, o tempo que decorre desde a produção do conhecimento até sua utilização por pesquisadores torna-se reduzido, como se observa a partir da dinâmica da maximização e aceleração do impacto da pesquisa, ilustrada na Figura 4.

O acesso aberto atende às demandas e interesses de diferentes atores que estão presentes no sistema de comunicação científica. Suas estratégias criam oportunidades que não existem no modelo tradicional de publicação científica. Suber (2010) explora essas oportunidades a cada um dos atores:

- a) autores: o acesso aberto promove uma audiência mundial, maior do que a possibilitada a qualquer periódico baseado em assinaturas, a despeito do grau de prestígio ou popularidade, e, comprovadamente, aumenta a visibilidade e o impacto do seu trabalho;
- b) leitores: o acesso aberto promove acesso livre de barreiras à literatura de que necessitam para realizar suas pesquisas, sem restrições pelo orçamento de bibliotecas onde podem ter privilégios de acesso. O acesso aberto aumenta o poder de recuperação e de alcance dos leitores. A literatura *online* livre pode ser entendida também como dados livres *online* para ferramentas que facilitam a busca no texto completo, indexação, mineração, resumo, tradução, consulta, criação de links, recomendação, alerta e outras formas de processamento e análise;
- c) professores e estudantes: a eliminação da necessidade de pagamentos ou permissões para reproduzir e distribuir conteúdo, diminui as desigualda-

- des de acesso a alguns recursos necessários para as pesquisas e ensino, problema que afeta especialmente os professores, alunos e pesquisadores de países subdesenvolvidos ou em desenvolvimento;
- d) bibliotecas: o acesso aberto resolve a crise dos periódicos científicos, também o que se chama de crise dos periódicos. Além disso, serve a outros interesses das bibliotecas, indiretamente. Bibliotecários querem ajudar seus usuários a encontrar informações que necessitam, independentemente dos limites orçamentários da biblioteca e de suas próprias coleções. Bibliotecários acadêmicos querem ajudar seus pesquisadores e instituição a aumentar sua audiência e impacto e ajudar a instituição;
  - e) universidades: o acesso aberto aumenta a visibilidade de seus pesquisadores e de suas pesquisas, reduz os gastos com assinaturas de periódicos e avança em sua missão de compartilhar o conhecimento;
  - f) periódicos e editores: o acesso aberto torna seus artigos mais visíveis, encontráveis, recuperáveis e utilizáveis. Se um periódico é de acesso aberto, ele pode usar dessa visibilidade superior para atrair submissões, publicidade, leitores e citações. Caso um periódico científico baseado em assinaturas promova acesso aberto a alguns de seus conteúdos (por exemplo, artigos selecionados em determinados números, números anteriores após um determinado período), ele pode usar disso para aumentar a visibilidade para atrair todos os mesmos benefícios das assinaturas. Por outro lado, caso um periódico permita o acesso aberto por meio do arquivamento em repositórios, isso se torna uma vantagem na atração de autores sobre aqueles periódicos que não permitem o arquivamento;
  - g) agências de fomento: o acesso aberto aumenta o retorno de seus investimentos em pesquisa, tornando os resultados das pesquisas financiadas mais amplamente disponíveis, mais encontráveis, mais recuperáveis e mais utilizáveis.

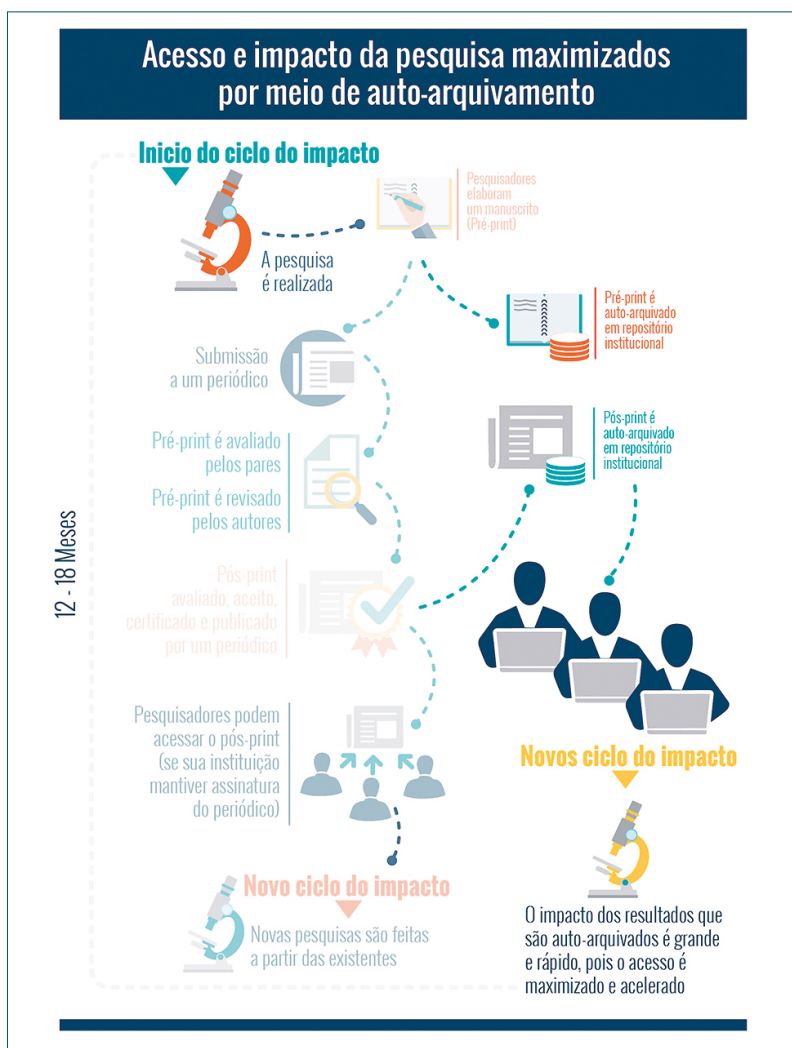


Figura 4 – Acesso e impacto da pesquisa maximizados por meio do autoarquivamento

Fonte: Adaptado de Brody e Harnad (2004).

Segundo a Declaração de Berlim (BERLIN DECLARATION ON OPEN ACCESS TO KNOWLEDGE IN THE SCIENCES AND HUMANITIES, 2003), o estabelecimento do acesso aberto como um procedimento vantajoso requer o empenho ativo de todo e qualquer indivíduo que produza conhecimento científico. Dessa maneira, as contribuições em acesso aberto podem incluir resultados de pesquisas científicas originais, dados de pesquisas não processados, metadados, fontes originais, representações digitais de materiais pictóricos, gráficos e material acadêmico multimídia. Além disso, devem satisfazer duas condições:

- a) autores e detentores dos direitos de tais contribuições concedem a todos os usuários o seguinte: direito gratuito, irrevogável e irrestrito de acessá-las; licença para copiá-las, usá-las, distribuí-las, transmiti-las e exibi-las publicamente; licença para realizar e distribuir obras derivadas, em qualquer suporte digital e para qualquer propósito responsável, em obediência à correta atribuição da autoria (as regras da comunidade continuarão a fornecer mecanismos para impor a atribuição e uso responsável dos trabalhos publicados, como acontece no presente) e com a garantia de fazer cópias;
- b) uma versão completa da obra e todos os materiais suplementares, incluindo uma cópia da licença, nos termos acima definidos, são depositados e, portanto, distribuídos em formato eletrônico normalizado e apropriado, em pelo menos um repositório que utilize normas técnicas adequadas (como as definições estabelecidas pelo modelo *Open Archives*) e que seja mantido por uma instituição acadêmica, sociedade científica, organismo governamental, ou outra organização estabelecida que pretenda promover o acesso aberto, a distribuição irrestrita, a interoperabilidade e o arquivamento a longo prazo.

Vê-se, portanto, que a reação da comunidade científica parte da convergência de soluções tecnológicas inovadoras, metodologias e o esforço do convencimento político em várias instâncias, constituindo-se em uma filosofia aberta, mencionada por Costa (2006). A filosofia aberta, segundo a autora, refere-se ao movimento observado nos últimos anos em direção ao uso de ferramentas, estratégias e metodologias que denotam novo modelo de representar um igualmente novo processo de comunicação científica, ao mesmo tempo que serve de base para interpretá-lo, compreendendo entre outras questões:

- a) software livre, para o desenvolvimento de aplicações em computador;
- b) arquivos abertos, para interoperabilidade em nível global;
- c) acesso aberto, para a disseminação ampla e irrestrita de resultados da pesquisa científica.

É fundamental, portanto, considerar o que observa Alberts (2002), ao afirmar que a informação científica e técnica é um bem público global, que deve estar livremente disponível para o benefício de todos. Weitzel (2006) considera que o movimento de acesso aberto vem construindo as condições necessárias para permitir o acesso irrestrito à produção científica legítima, alterando não somente o processo de aquisição de informação científica, mas também a sua produção, disseminação e uso. E, nesse contexto, conforme Johnson (2002), cresce claramente o papel de modelos alternativos de comunicação científica, tais como repositórios institucionais, ao quebrarem monopólios de editores e promoverem a preocupação de pesquisadores sobre a produção intelectual das universidades

e institutos de pesquisa. A Budapest Open Access Initiative (2001) recomendou duas estratégias complementares para que de fato a literatura científica esteja disponível e acessível:

- a) a via dourada, que significa o acesso aberto promovido nos próprios periódicos científicos, de modo que os artigos científicos possam ser disseminados sem restrições de acesso ou uso;
- b) a via verde, que significa o sinal verde de editores científicos para o autoarquivamento da produção científica pelos próprios autores em repositórios digitais de acesso aberto, especialmente em repositórios institucionais.

Segundo Suber (2010), a principal diferença entre as duas estratégias reside no fato de que periódicos de acesso aberto conduzem o processo de avaliação por pares e os repositórios digitais não. Para o autor as diferenças estão relacionadas também com os locais e veículos de entrega da informação, e não com direitos dos usuários ou mesmo com o grau de abertura.

## **ACESSO ABERTO NO BRASIL: UM MODELO *EX-POST FACTO***

O desenvolvimento do acesso aberto no Brasil foi em grande parte resultado de esforços do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), muitas vezes em parceria com universidades, como foi o caso bem-sucedido com a Universidade de Brasília (UnB).

Tendo em vista sua missão, que é promover a competência, o desenvolvimento de recursos e a infraestrutura de informação científica e tecnológica para a produção, a socialização e a integração do conhecimento científico-tecnológico, o Ibict investiu esforços visando à operacionalização do acesso aberto no Brasil. Ressalta-se que o percurso apresentado a seguir consiste em uma leitura *ex-post facto* da implementação do acesso aberto no Brasil, ou seja, é uma tentativa de consubstanciar em um modelo as ações brasileiras conduzidas principalmente pelo Ibict.

Em sintonia com os desenvolvimentos mundiais, o Ibict estabeleceu como estratégias básicas as orientações da via dourada e da via verde para o acesso aberto à informação científica. Após uma análise detalhada das ações empreendidas, pôde-se enxergar o que se pode denominar de modelo brasileiro. Nessa perspectiva, informação científica foi compreendida como todas as manifestações da literatura científica. Como linhas de ação, foram estabelecidas quatro dimensões básicas inter-relacionadas (Figura 5): capacitação, tecnologia, sistemas de informação e políticas.



Figura 5 – Modelo norteador das ações de acesso aberto do Ibict

Fonte: Autoria própria (2017).

Ou seja, todas as ações de acesso aberto promovidas pelo Ibict podem ser agrupadas nas duas grandes estratégias (via dourada e via verde), as quais, por sua vez, foram viabilizadas por meio de ações agrupadas em quatro dimensões. A primeira delas foi o desenvolvimento de competências e habilidades relacionadas com o acesso aberto e comunicação científica nos níveis institucional (do próprio Ibict) e nacional (junto às bibliotecas de universidades e institutos de pesquisa e editores científicos) a partir de ações de capacitação. A segunda foram as ações referentes à tecnologia e disseram respeito à adaptação, aperfeiçoamento e transferência de tecnologias para a construção de uma infraestrutura tecnológica nacional para a instauração das duas estratégias de acesso aberto. As ações agrupadas na terceira dimensão denominada sistemas de informação contemplaram, por sua vez, a construção e o suporte à construção de sistemas de informação de acesso aberto: repositórios institucionais (RI), periódicos científicos eletrônicos, bibliotecas digitais de teses e dissertações, mecanismos de busca e sistemas de informação de apoio aos gestores. Por fim, as ações compreendidas na quarta dimensão corresponderam à atuação política, traduzida em um conjunto de subações cujos objetivos foram sensibilizar e influenciar comportamento dos atores envolvidos no cenário do acesso aberto. A seguir cada uma das linhas de ação é especificada em função das atividades realizadas no âmbito das ações de acesso aberto do Ibict.

## CAPACITAÇÃO

A operacionalização das estratégias de acesso aberto no Brasil requereu o desenvolvimento de competências e habilidades relacionadas com o acesso aberto e comunicação científica tanto em relação à constituição de capacidades institucionais da equipe do próprio Ibict quanto, principalmente, em nível nacional, ou seja, o treinamento de equipes gestoras de repositórios digitais (bibliotecas digitais de teses e dissertações, repositórios institucionais) e periódicos científicos eletrônicos de acesso aberto. No que tange aos repositórios digitais, a capacitação envolveu basicamente o treinamento de bibliotecários e analistas de sistemas de universidades e institutos de pesquisa brasileiros para o seu planejamento e implementação.

Bibliotecários foram sistematicamente expostos ao contexto da comunicação científica e do acesso aberto, bem como melhores práticas para o planejamento, implementação e funcionamento de repositórios institucionais, e, por último às ferramentas TEDE e DSpace. Analistas de sistemas, por seu turno, foram submetidos a treinamento de instalação e configuração destas mesmas ferramentas. Do ponto de vista formal, o desenvolvimento de competências e habilidades ocorreu a partir da realização dos seguintes eventos:

- a) treinamentos presenciais sobre o uso do TEDE para a construção de bibliotecas digitais de teses e dissertações nas instituições;
- b) treinamentos presenciais sobre o uso do SEER (foram ministrados 59 treinamentos presenciais em todo o Brasil com cerca de 400 participantes);
- c) treinamentos a distância sobre o uso do SEERaD (10 turmas);
- d) realização do I e II Encontro de Usuários do SEER (EUSEER);
- e) realização de três Workshops para a Construção de Repositórios Institucionais e um Encontro sobre Gestão de Repositórios Institucionais;
- f) eventos sobre acesso aberto promovidos pelo Ibict (10 eventos);
- g) elaboração e distribuição de conteúdos para capacitação (livro, cartilha e manuais).

Uma quantidade considerável de ações de capacitação foi realizada em forma de prestação de assessorias, oficinas, cursos e participação em fóruns para repasse da experiência e formação para a criação de RI e periódicos eletrônicos. Além disso, prestou-se suporte direto para implantação e gerenciamento de sistemas de informação de acesso aberto em inúmeras instituições em todo o país.

## TECNOLOGIA

As ações relacionadas a esta dimensão tiveram por objetivo promover a absorção de tecnologias de informação e de comunicação aplicadas ao tratamento e disseminação da informação científica. A linha de ação foi responsável pela instauração da infraestrutura tecnológica necessária à operacionalização da via dourada e da via verde para o acesso aberto no Brasil. As ações foram conduzidas em dois estágios.

O primeiro estágio foi relacionado com a prospecção, construção/adaptação de ferramentas para a criação de repositórios digitais (bibliotecas digitais de teses e dissertações, repositórios institucionais, repositórios temáticos) e periódicos científicos eletrônicos de acesso aberto. Além da análise e testes de outras ferramentas, o Ibict definiu as seguintes como prioritárias:

- a) Sistema de Publicação Eletrônica de Teses e Dissertações (TEDE), para a construção e gerenciamento de bibliotecas digitais de teses e dissertações (software desenvolvido pelo próprio instituto a partir de padrões internacionais);
- b) Sistema Eletrônico de Editoração de Revistas (SEER), para a criação de portais e periódicos científicos eletrônicos (tradução e adaptação do software *Open Journal System – OJS*);
- c) DSpace, para a construção de repositórios digitais (tradução, adaptação e customização);
- d) Sistema *Online* de Acompanhamento de Conferências (SOAC), para a construção de portais e conferências científicas eletrônicas (tradução do software *Open Conference System – OCS*).

Outros softwares também constituíram o conjunto de ferramentas do programa Acesso Aberto do Ibict, porém, com menor ênfase:

- a) *Eprints*, denominado no Brasil Diálogos Científicos (DiCi), para a construção de repositórios digitais;
- b) Bibliotecas Virtuais Temáticas (BVT), cujo objetivo é facilitar o acesso à informação pelo usuário a diferentes fontes de informação, sendo totalmente flexível na sua construção, permitindo a cada instituição customizar de acordo com as suas características e necessidades;
- c) Tecer, software para a gestão de linguagens documentárias.

O segundo estágio desta linha de ação esteve relacionado com a transferência das tecnologias estabelecidas como prioritárias para a promoção do acesso aberto. Além de sua relação direta com a dimensão capacitação, onde o conhecimento a respeito do funcionamento e gestão das ferramentas era compartilhado com a comunidade (gestores de repositórios digitais e de portais e periódicos



científicos eletrônicos), a transferência das tecnologias dependeu também da construção de ambientes de referência em que a comunidade pudesse se desenvolver e colaborar mutuamente. Foram eles:

- a) serviço de Incubadora de Revistas Científicas criadas com o uso do SEER (INSEER), cujo objetivo é apoiar e estimular a construção e manutenção de periódicos científicos de acesso aberto na Internet;
- b) desenvolvimento de página contendo informações sobre o DSpace onde são encontradas informações sobre os repositórios criados com o software no Brasil, documentos úteis relacionados com o DSpace, perguntas frequentes, link para efetuar o download gratuito e outros links relacionados com a ferramenta;
- c) desenvolvimento do Portal do SEER contendo informações sobre revistas criadas, portais de revistas, treinamentos, link para download, documentos úteis relacionados com a operação do sistema, perguntas frequentes e links relacionados.

Além das ações descritas, em determinado momento o Ibict constatou que as universidades e institutos de pesquisa, em sua totalidade instituições públicas, eram deficientes em suas infraestruturas de hardware para a construção de seus respectivos provedores de dados. Tendo em vista isso, o instituto adquiriu e distribuiu kits tecnológicos (servidores com os sistemas SEER e DSpace devidamente instalados e configurados) de modo a dotá-las com o hardware que possibilitasse a construção de seus repositórios institucionais e suas revistas científicas.

## SISTEMAS DE INFORMAÇÃO

A via dourada e a via verde, estratégias constituintes do modelo norteador das ações de acesso aberto do Ibict, pressupõem a construção de sistemas de informação de acesso aberto à informação científica. Nessa perspectiva, adotou-se o entendimento de que sistema de informação é a combinação organizada de pessoas, hardware, software, redes de comunicação, recursos de informação, políticas e procedimentos que visam reunir, organizar, armazenar, preservar, recuperar e disseminar informação que alimenta e que resulta de atividades de pesquisa. No âmbito das ações do instituto, os sistemas de informação são essencialmente de dois tipos, a saber:

- a) sistemas de informação provedores de dados: são os repositórios digitais (bibliotecas digitais de teses e dissertações, repositórios institucionais) e os periódicos científicos eletrônicos de acesso aberto, todos baseados no protocolo OAI-PMH. O Ibict participou diretamente da construção de:
  - 105 bibliotecas digitais de teses e dissertações;

- 43 repositórios institucionais criados com suporte direto do Ibict;
  - cerca de 1.600 periódicos científicos eletrônicos;
- b) sistemas de informação provedores de serviços: são os mecanismos de busca construídos para integrar os provedores de dados. Por intermédio ou por meio de participação do Ibict, o Brasil conta com os seguintes provedores de serviços:
- Biblioteca Digital Brasileira de Teses e Dissertações (BDTD): o Ibict coordena o projeto BDTD, que integra os sistemas de informação de teses e dissertações existentes nas instituições de ensino e pesquisa brasileiras, e também estimula o registro e a publicação de teses e dissertações em meio eletrônico;
  - Portal Brasileiro de Acesso Aberto à Informação Científica (oasisbr): é um mecanismo de busca multidisciplinar que permite o acesso gratuito à produção científica de autores vinculados a universidades e institutos de pesquisa brasileiros. Por meio do oasisbr é possível também realizar buscas em fontes de informação portuguesas;
  - Diretório Luso-Brasileiro: o sistema pretende agregar as fontes de informação científica disponíveis em acesso aberto em Portugal e no Brasil, permitindo-lhes ter mais visibilidade e destaque.

Além desses dois tipos de sistemas de informação (provedores de dados e provedores de serviços), o Ibict implantou o Diadorim, um sistema de informação de apoio à gestão de repositórios digitais cujo objetivo é identificar, sistematizar e disponibilizar as informações das políticas estabelecidas pelas editoras dos periódicos brasileiros referentes ao armazenamento dos artigos nos repositórios institucionais.

## POLÍTICAS

A dimensão denominada Políticas ocupa uma relevante posição. A atuação política pode ser traduzida em um conjunto de subações cujos objetivos foram sensibilizar e influenciar comportamento de atores envolvidos no cenário do acesso aberto, nomeadamente pesquisadores, governo, bibliotecas, universidades e institutos de pesquisa, agências de fomento e editores científicos. Nesse sentido, o Ibict induziu e participou de um grupo de ações políticas, entre as quais as mais relevantes foram:

- a) lançamento do Manifesto Brasileiro de Apoio ao Acesso Livre (2005);
- b) lançamento e assinatura da Declaração de Florianópolis (2006);
- c) assinatura da Declaração de Berlim (2006);

- d) Projeto de Lei 1120/2007<sup>1</sup>, que dispõe sobre o processo de disseminação da produção técnico-científica pelas instituições de ensino superior no Brasil e dá outras providências. O projeto – que obrigava as instituições públicas de ensino superior a construir os repositórios institucionais para depósito do inteiro teor da produção técnico-científica do corpo discente e docente – foi arquivado e encerrou sua tramitação no início do ano de 2012;
- e) Projeto de Lei do Senado 387/2011<sup>2</sup>, que torna compulsório as instituições de educação superior de caráter público, bem como as unidades de pesquisa, a construir repositórios institucionais de acesso livre, nos quais deverão ser depositados o inteiro teor da produção técnico-científica conclusiva dos estudantes aprovados em cursos de mestrado, doutorado, pós-doutorado ou similar, assim como da produção técnico-científica, resultado de pesquisas científicas realizadas por professores, pesquisadores e colaboradores, apoiados com recursos públicos para acesso livre na rede mundial de computadores; entende-se por produção técnico-científica monografias, teses, dissertações e artigos publicados em revistas, nacionais e internacionais, com revisão por pares;
- f) articulação junto aos diversos segmentos da comunidade científica visando sensibilizá-los quanto à importância do acesso aberto (sociedades científicas, associações de bibliotecários, associação de editores científicos, associações de gestores e dirigentes universitários, reitores de universidades, dirigentes de agências de fomento, entre outros).

Além dessas ações específicas, o Ibict atuou diretamente sobre universidades e institutos de pesquisa no que concerne ao estabelecimento de suas políticas institucionais de acesso aberto, sendo as políticas de depósito compulsório as mais pautadas. Nessa perspectiva o instituto, como unidade de pesquisa vinculada ao Ministério da Ciência, Tecnologia e Inovação, estabeleceu sua própria política mandatória de acesso aberto no ano de 2009.

## **ACESSO ABERTO X GESTÃO DA INFORMAÇÃO CIENTÍFICA: DESCORTINANDO RELAÇÕES**

É imprescindível que qualquer discussão a respeito do acesso aberto ou de gestão da informação científica não esteja dissociada das noções e do contexto da comunicação científica. Em última análise, a finalidade de ambas as práticas é promover fluxos de informação científica que atendam e tornem mais fluida a comunicação científica, atividade inexorável da pesquisa e produção do conhecimento. Compreende-se, portanto, que o acesso aberto existe para aperfeiçoar

1 Disponível em: <<http://www.camara.gov.br/sileg/integras/461698.pdf>>. Acesso em: 24 jul. 2017.

2 Disponível em: <<http://legis.senado.leg.br/sdleg-getter/documento?dm=2949201&disposition=inline>>. Acesso em: 24 jul. 2017.

a comunicação científica e, para tanto, lança mão de processos de gestão da informação, conforme ilustrado na Figura 6.

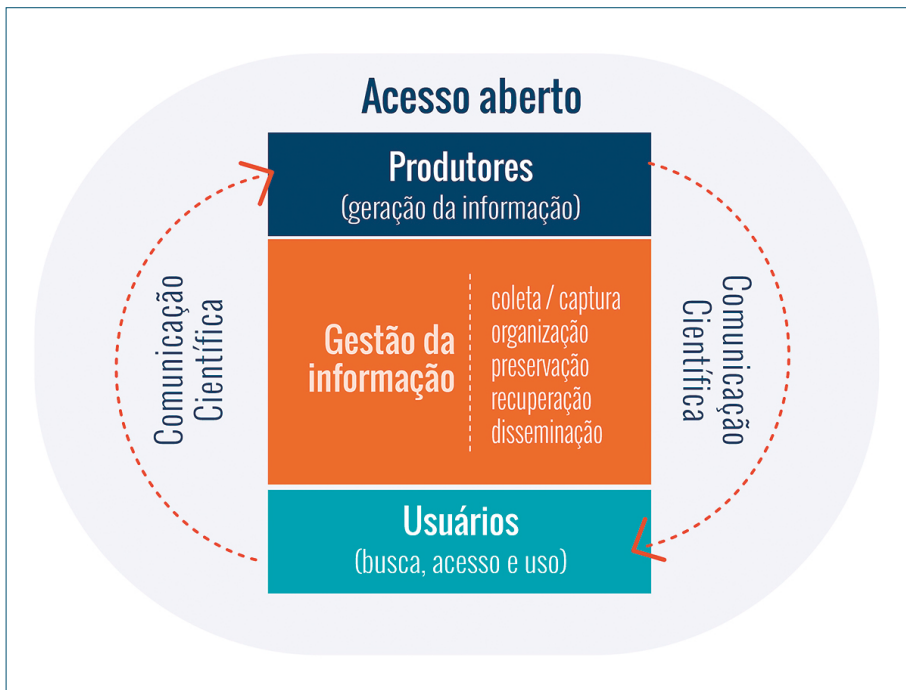


Figura 6 – Comunicação científica, acesso aberto e gestão da informação científica  
Fonte: Autoria própria (2017).

Os fluxos desimpedidos de informação entre produtores e usuários da informação – funções intercambiáveis e dependentes do estágio em que se encontram as atividades de pesquisa – dependem, necessariamente, de processos de gestão da informação. Uma vez orientados pelos pressupostos do acesso aberto, tais processos, atividades típicas de sistemas de informação em uma abordagem ampla, viabilizam o fluxo da informação. Ou seja, o fluxo desimpedido da informação entre produtores e usuários pressupõe que a informação seja coletada/reunida, organizada, preservada, tornada recuperável e disseminada segundo requisitos do acesso aberto.

A relação entre as três práticas foi objeto de pesquisa de Leite (2011), cujo principal resultado foi a proposição de um modelo de gestão da informação científica cujo detalhamento encontra-se em Leite (2011) e Leite e Costa (2016).

## UM BREVE APONTAMENTO DE AVANÇOS, OPORTUNIDADES E DESAFIOS

As ações relatadas refletem o estágio do desenvolvimento do acesso aberto na época em que foram implantadas. Percebe-se que, ao longo do período, que os resultados da implantação das ações foram bem-sucedidos em determinados aspectos, inócuos em outros. A seguir são apontados alguns avanços, oportunidades e desafios.

Costa e Leite (2017) consideram que, em razão de sua desestruturação histórica associada a outros fatores, o sistema de comunicação científica da América Latina e Caribe não foi acompanhado pelo progresso de uma indústria, a exemplo da Europa e Estados Unidos. Os autores querem dizer com isso que, por um lado, durante muito tempo, este fato contribuiu para invisibilidade da ciência produzida na região, mas, por outro, mais recentemente, foi exatamente o fator que contribuiu para a bem-sucedida estratégia de ampla comunicação científica por meio de periódicos de acesso aberto. Os autores pontuam, no entanto, que os avanços nesse campo devem também ser avaliados em razão da qualidade dos periódicos. A conclusão de Costa e Leite (2017) evidencia o protagonismo brasileiro no que se refere aos avanços em ciência e tecnologia (C&T), e sua comunicação, na região. Os indicadores demonstram que o país responde por quantidade significativa dos resultados de pesquisa que são disseminados internacionalmente.

Até o ano de 2017 cerca de 6 mil periódicos científicos de acesso aberto já haviam sido criados no Brasil, segundo dados disponibilizados pelo Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal (Latindex)<sup>3</sup>. No entanto, destes, apenas 128 encontram-se indexados na *Web of Science*<sup>4</sup>, 334 indexados na Scopus<sup>5</sup> e 285 na SciELO<sup>6</sup>. Isso quer dizer que o Brasil avançou em termos da quantidade de periódicos científicos de acesso aberto, mas ainda precisa equilibrar a relação entre quantidade e qualidade dos periódicos. Inequivocamente, no que diz respeito à qualidade de periódicos científicos de acesso aberto, a principal iniciativa brasileira é a SciELO.

A via verde, por sua vez, também galgou avanços no Brasil. Contudo, necessita superar desafios e aproveitar oportunidades que se apresentam tanto no Brasil quanto na América Latina e Caribe. Como resultado das ações implantadas ao longo do tempo, pode-se afirmar:

3 Disponível em: <<http://www.latindex.org/latindex/tablaPais?id=9&id2=0>>. Acesso em: 09 ago. 2017.

4 Disponível em: <<https://www.webofknowledge.com/>>. Acesso em: 09 ago. 2017.

5 Disponível em: <<https://www.scopus.com/>>. Acesso em: 09 ago. 2017.

6 Disponível em: <<http://www.scielo.br/>>. Acesso em: 09 ago. 2017.

- a) a questão tecnológica parece não constituir mais um entrave: uso predominante do DSpace;
- b) boa parte das universidades e instituições de pesquisa estão inseridas ou atentas ao acesso aberto;
- c) a BDTD encontra-se em alto grau de maturidade e plenamente consolidada. Poucos países no mundo atingiram tamanha quantidade de teses e dissertações acessíveis na Internet. Segundo os dados disponíveis no portal da *Networked Digital Library of Theses and Dissertations* (NDLTD)<sup>7</sup>, a BDTD disponibiliza cerca de 400 mil documentos no sistema, sendo, portanto, a quarta maior provedora de dados para o Portal;
- d) repositórios institucionais brasileiros são quase inteiramente gerenciados por bibliotecas de universidades ou institutos de pesquisa. Tal fato favorece a qualidade das operações desses sistemas;
- e) parece não haver qualquer relação entre as facilidades da via dourada e a propulsão da via verde. Ou seja, a literatura científica publicada no Brasil, assim como na América Latina e Caribe, é dissociada de interesses comerciais, mas, ainda assim, não se encontra em repositórios institucionais;
- f) maior parte dos repositórios institucionais brasileiros enfrentam problemas de povoamento. Segundo o *Ranking Web of Repositories*<sup>8</sup>, os cinco sistemas brasileiros dessa natureza melhor pontuados são originários das seguintes instituições<sup>9</sup>: Universidade Federal do Rio Grande do Sul (UFRGS): cerca de 155 mil documentos, Universidade de Campinas (UNICAMP): cerca de 68 mil documentos, Universidade Estadual Paulista (UNESP): cerca de 111 mil documentos, Universidade Federal de Santa Catarina (UFSC): cerca de 95 mil documentos, Empresa Brasileira de Pesquisa Agropecuária (Embrapa): cerca de 80 mil documentos;
- g) parte dos repositórios institucionais brasileiros gerenciam conteúdos de natureza distinta da literatura científica e incluem documentação administrativa e de memória organizacional da instituição. Esse fenômeno pode ter ocorrido em razão da não compreensão e consequente distorção da finalidade maior do acesso aberto e da comunicação científica, ou simplesmente por terem cedido às pressões para solução de problemas de gestão da informação de outra natureza na instituição;
- h) mandatos de depósito compulsório constituem uma das molas propulsoras do povoamento de repositórios institucionais. No entanto, tornam-se inócuos se não considerarem determinados aspectos ou se dissociados de outras ações. Segundo dados do diretório *Policies Melibea*<sup>10</sup>, foram registrados apenas 11 documentos brasileiros com caráter de po-

7 Dados de maio de 2017. Disponível em: <<http://search.ndltd.org/>>. Acesso em: 24 jul. 2017.

8 Disponível em: <[http://repositories.webometrics.info/en/Latin\\_America/Brazil](http://repositories.webometrics.info/en/Latin_America/Brazil)>. Acesso em: 03 maio 2017.

9 Dados de maio de 2017.

10 Policies Melibea: directory and estimator policies for open access to scientific production. Disponível em: <<http://www.accesoabierto.net/politicas/>>. Acesso em: 03 maio 2017.

lítica para o acesso aberto. Embora os documentos sejam conquistas importantes para a promoção do acesso aberto em suas instituições, via de regra, eles não determinam o depósito compulsório, não estabelecem prazo para o depósito dos documentos e não especificam as sanções em caso de descumprimento dos termos estabelecidos. A obrigatoriedade do depósito é um aspecto fundamental para o fortalecimento da política como instrumento promotor do acesso aberto. No entanto, pondera-se que o estabelecimento da obrigatoriedade não significa o seu cumprimento, sendo necessário o seu atrelamento a outros processos, como, por exemplo, progressão na carreira de docente-pesquisador, avaliação de programas de pós-graduação, distribuição de recursos entre unidades acadêmicas nas universidades ou institutos de pesquisa, presença da obrigatoriedade em editais internos de fomento às atividades de pesquisa e participação em eventos, entre outros. Além disso, mandatos de depósito compulsório devem ser acompanhados de ações de marketing e sensibilização;

- i) como em todo o mundo, o povoamento de repositórios institucionais é um desafio. A expectativa do autoarquivamento em sentido estrito foi, ao longo do tempo, reconhecida como inviável, dando lugar a estratégias alternativas de depósito, entre elas o depósito automático. Em linhas gerais, significa a importação de lotes de metadados e de arquivos relacionados e sua inserção automática no repositório, o que permite o aumento considerável de itens depositados. Possivelmente por limitações tecnológicas, maior parte dos repositórios brasileiros não adotam essa estratégia.

Os autores deste capítulo consideram que a força motriz para a alavancagem do acesso aberto no Brasil, a exemplo de outros países bem-sucedidos, tem origem nas agências de fomento. A despeito da importância da rede de agências de fomento existente no Brasil para o acesso aberto, constituída pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e fundações estaduais de amparo à pesquisa, o maior poder de estímulo ao acesso aberto está nas mãos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). A agência financia os programas de pós-graduação no Brasil, celeiros de maior parte da pesquisa científica realizada no país. Além do financiamento, a CAPES é responsável por avaliar todos os programas de pós-graduação brasileiros, de modo que seja assegurada e mantida a qualidade dos cursos de mestrado e de doutorado no país. Por deter prerrogativa exclusiva de avaliar e validar cursos de mestrado e doutorado, a CAPES possui a legitimidade necessária para estabelecer que os programas de pós-graduação tenham sua produção científica depositada em repositórios institucionais. Uma eventual determinação dessa natureza, além de desejável, seria facilitada em razão do fato de que a Plataforma Sucupira, sistema de informação adotada no processo de avaliação, registra a totalidade da produção científica dos cursos de mestrado e doutorado.

## AÇÕES FUTURAS

Tendo em vista os possíveis desdobramentos das ações descritas em um Programa de Acesso Aberto, torna-se imperativo o estabelecimento de novos rumos de modo a fomentar o desenvolvimento da infraestrutura de informação com vistas a promover avanços no sistema científico e tecnológico do país. A primeira recomendação é que o Ibict repense dois conceitos fundamentais a partir dos quais todas as ações de acesso aberto foram fundadas.

O primeiro diz respeito ao conceito de informação científica. No âmbito das ações que se seguiram, informação científica foi tratada como sinônimo de literatura científica. Esta, por sua vez, diz respeito a todos os tipos de documentos científicos gerados ao longo da condução das atividades de pesquisa. Nessa perspectiva, informação científica compreende relatórios científicos, comunicações publicadas em anais de eventos, teses e dissertações, livros, artigos de periódicos científicos e capítulos de livros, além de outras poucas variações que representam peculiaridades de padrões de comunicação de áreas. Entretanto, de acordo com a definição da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO) (AGUIAR, 1991, p. 8), adotada pelo Ibict ao longo das décadas de 1970 e 1980, informação científica e técnica diz respeito a:

Elementos simbólicos utilizados para comunicar o conhecimento científico e técnico, independente de seu caráter (numérico, textual, icônico etc.), dos suportes materiais, da forma de apresentação. Refere-se tanto à substância ou conteúdo dos documentos quanto à sua existência material. Também se emprega esse termo para designar tanto a mensagem (conteúdo e forma) quanto sua comunicação (ação). Quando necessário, distingue-se entre informação bruta (fatos, conceitos, representações) e os documentos, em que se acha registrada.

Muito embora date de cerca de quatro décadas, sua configuração engloba outros tantos elementos que vão além da literatura científica, permitindo incluir outras manifestações informacionais do conhecimento científico e técnico como, por exemplo, dados brutos de pesquisa e imagens. É importante perceber que, ainda no ano de 1971, estas perspectivas já se faziam necessárias e, certamente, nos dias de hoje, frente às possibilidades e facilidades tecnológicas para a produção de informação digital, elas o são ainda mais.

O segundo conceito objeto de reflexão por parte do Instituto é a ampliação do escopo do sistema de comunicação em que as suas ações estão inseridas. Tradicionalmente, o Ibict desenvolve suas atividades voltadas para o contexto da comunicação científica, ou seja, de fluxos de informação entre pesquisadores. Face à crescente relevância de aumento de impacto social das atividades científicas e



tecnológicas, é fundamental que sejam constituídas infraestruturas de informação que promovam diálogos entre ciência, tecnologia e sociedade. Nesse sentido, ações de divulgação científica constituem esforços nobres, complementares e expansivos da comunicação científica. Assim, do mesmo modo que a comunicação científica, ações de divulgação científica dependerão de processos de gestão da informação que sistematizem os fluxos de informação entre produtores de informação (pesquisadores) e usuários da informação (sociedade). Processos de gestão da informação neste ambiente devem se beneficiar dos avanços já galgados pelo movimento mundial em favor de acesso aberto.

## DELINEAMENTO GENÉRICO DE UM PROGRAMA DE GESTÃO DA INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA DE ACESSO ABERTO PARA O BRASIL

Promover a competência, desenvolvimento de recursos e infraestrutura de informação científica e tecnológica para a produção, a socialização e a integração do conhecimento científico-tecnológico significa, em última análise, criar condições apropriadas para que ocorram fluxos de informação desimpedidos entre produtores e usuários de conhecimento resultante das atividades de pesquisa, desenvolvimento e inovação. Ou seja, o foco está na efetiva comunicação da informação entre quem cria e quem utiliza informação. Embora isso seja fundamental, investimentos em infraestrutura de informação e comunicação científica *per se* já não são suficientes para lidar com a complexidade dos desafios informacionais emergentes. Além de repensar estratégias para lidar com a informação científica e tecnológica na perspectiva de produtores e usuários, é necessário que isso seja feito de modo coordenado e planejado, considerando as transformações em curso e a complexidade do sistema global de informação e comunicação científica. Para tanto, a implementação de estratégias sistematizadas de gestão da informação científica e tecnológica que levem em conta a existência de forças que influenciam a produção, disseminação e uso da informação e do conhecimento no contexto científico e tecnológico. Quer se dizer com isso que a efetiva comunicação entre produtores e usuários de informação depende da implantação de processos sistematizados de gestão da informação científica e tecnológica.

Partindo dos pressupostos acima, apresenta-se uma proposta de modelo norteador para um eventual programa de Gestão da Informação Científica e Tecnológica de Acesso Aberto, para o Ibict (Figura 7).

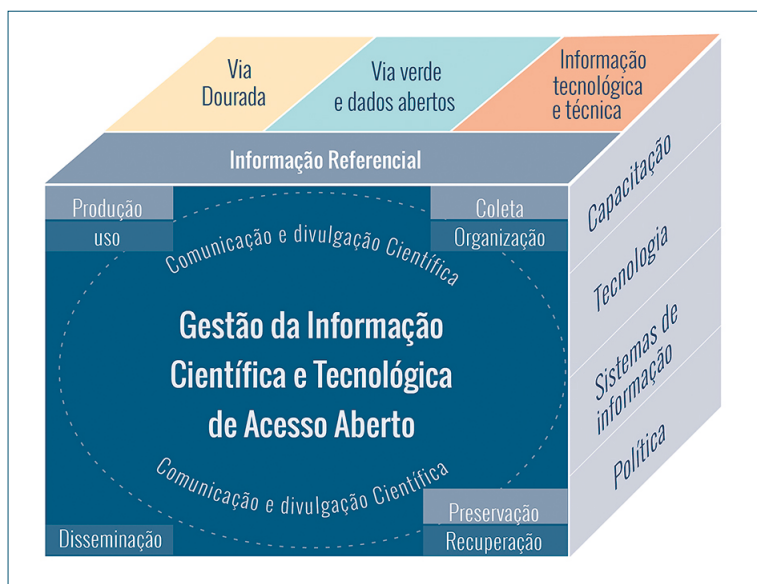


Figura 7 – Modelo preliminar de um programa de Gestão da Informação Científica e Tecnológica de Acesso Aberto para o Ibict  
Fonte: Autoria própria (2017).

A proposta de modelo é baseada em práticas até então adotadas pelo instituto, Figura 3, com algumas mudanças fundamentais e, em Leite (2011). O detalhamento do modelo:

- o aspecto central tornou-se o conjunto de processos que constituem a gestão da informação científica e tecnológica de acesso aberto e não mais apenas o acesso aberto por si. Ou seja, o foco está na implantação de um programa de gestão da informação de modo que seja possível, diante dos desafios informacionais contemporâneos, sistematizar os fluxos de informação que alimentam e que resultam das atividades de pesquisa, desenvolvimento e inovação;
- as estratégias básicas e tradicionais de acesso aberto foram ampliadas de modo que a via verde contemple também a noção de ciência aberta, especialmente o ciclo de gestão de dados brutos de pesquisa. Uma terceira estratégia é sugerida de modo que a informação tecnológica/técnica possa ser incluída ao ponto em a dimensão da divulgação científica possa ser incorporada como uma ampliação do contexto da comunicação científica em que se insere o instituto;
- à via dourada, via verde/dados abertos e informação tecnológica inclui-se uma camada transversal às três estratégias: Informação Referencial. Ou seja, o modelo prevê como necessária a integração entre fontes de informação de inteiro teor e fontes de informação referenciais, especialmente

catálogos de bibliotecas e centros de documentação de universidades e institutos de pesquisa. Esta camada transversal incorporada constitui também ponto de partida para o repensar de produtos e serviços de informação tradicionais do instituto como COMUT e CCN;

- d) orientação explícita a processos de gestão da informação, nomeadamente produção, coleta, organização, preservação, recuperação, disseminação e uso tanto da informação científica quanto da informação tecnológica/técnica em ambiente digital e impresso;
- e) no modelo anterior, o Ibict pautava-se em linhas de ação genéricas (capacitação, tecnologia, sistemas de informação e políticas) sem orientação estratégica à gestão da informação. Na nova proposta, tanto a comunicação (informação científica, incluindo dados abertos) quanto a divulgação científica (informação tecnológica/técnica) são promovidas apropriadamente em razão da sistematização de seus fluxos, que será alcançada a partir de processos de gestão da informação (Figura 8);



Figura 8 – Relação entre gestão da informação científica e tecnológica de acesso aberto, comunicação e divulgação da informação científica  
Fonte: Autoria própria (2017).

- f) o instituto permanece como o articulador, em nível nacional, dos processos informacionais das instituições cuja missão primordial é a produção do conhecimento em ciência e tecnologia. Ou seja, o Ibict deve propor e coordenar uma rede e sub-redes de informação as quais, por sua vez, são constituídas por sistemas de informação científica, tecnológica e técnica mantidos por instituições produtoras de conhecimento;
- g) a ideia de usuário no contexto do instituto é ampliada. De membros da comunidade científica que produzem e consomem informação científica mediante processos de comunicação científica, o Ibict incorpora sinergicamente em suas estratégias outras categorias de usuários, como exemplo, profissionais das diversas áreas, técnicos, tecnólogos, cidadãos comuns, entre outros;

- h) o modelo chama a atenção para a necessidade de o instituto qualificar suas ações de inclusão digital e social, vinculando-as a contextos específicos em que se inserirão ações de divulgação científica (saúde, agricultura, meio ambiente e ecologia, educação, mobilidade urbana, entre outros).

Além das direções estratégicas apontadas pelo modelo proposto, apontam-se a seguir linhas de ação que podem fazer parte de futuro Programa de Gestão da Informação Científica e Tecnológica Brasileiro:

- a) integração de sistemas nacionais de informação em ciência e tecnologia, incluindo sistemas de gestão, como é o caso da Plataforma Lattes do CNPq;
- b) qualificação e estratificação da produção científica brasileira de acesso aberto de modo que seja possível a oferta de outros produtos e serviços de informação com valor agregado;
- c) fusão entre BDTD, oasisbr e catálogos de bibliotecas de universidades e institutos de pesquisa brasileiros;
- d) potencializar a inclusão das ciências sociais e humanidades a partir da capacitação e difusão do Open Monograph Press (OMP), ferramenta do *Public Knowledge Project* (PKP) para a automatização de processos editoriais e publicação de livros em ambiente de acesso aberto;
- e) potencializar a difusão do SOAC para outras áreas do conhecimento que não a ciência da informação no Brasil, tendo em vista que a publicação em anais de conferências é tendência crescente em todas as áreas;
- f) atuar colaborativamente com infraestruturas nacionais de *e-science* como, por exemplo: Sistema Nacional de Processamento de Alto Desempenho (SINAPAD), Rede Galileu, GridUNESP e E-science grid facility for Europe and Latin America (EELA);
- g) propor, induzir e/ou apoiar a construção de sistemas de informação tecnológica voltados para os diferentes segmentos produtivos (desde indústria de grande porte a pequenos produtores);
- h) no que concerne a iniciativas de divulgação científica e gestão da informação tecnológica/técnica, priorizar áreas com alto impacto social como saúde, agricultura e meio ambiente;
- i) formar e coordenar a rede brasileira de informação científica e tecnológica de acesso aberto;
- j) realizar estudos métricos que permitam gerar conhecimento a partir dos bancos de dados mantidos ou que fazem parte de redes coordenadas pelo Instituto;
- k) adotar padrão de geração de indicadores de uso de sistemas de informação de acesso aberto;

- l) elaborar metodologia de **ranqueamento** de universidades e institutos de pesquisa brasileiros baseada na acessibilidade a seus conteúdos na internet.

## CONSIDERAÇÕES FINAIS

A despeito das oportunidades a aproveitar e desafios a superar, do ponto de vista das tradicionais estratégias via dourada e via verde, o acesso aberto no Brasil evoluiu o quanto pôde. Já se tem conhecimento sobre o que, como e onde é possível chegar. Contudo, com base na argumentação deste texto, é possível perceber que a atuação pautada na ótica tradicional do acesso aberto, contemplando simplesmente a capacitação, tecnologia, sistemas de informação e políticas, não é suficiente para avançar e inovar nos processos do sistema de comunicação científica.

O avanço e inovação nesse contexto dependem, sim, das quatro dimensões, mas, sobretudo, não prescindem da articulação de outras perspectivas. A primeira é a noção de que processos de gestão da informação científica orientados pelos pressupostos do acesso aberto são condições *sine qua non* para a melhoria da comunicação científica. A segunda é o investimento e incursão definitiva do país na conjuntura da ciência aberta. Dessa ótica, desdobram-se outras duas demandas emergentes: a gestão de dados brutos de pesquisa e ampliação do alcance dos avanços científicos para além da comunidade científica, decorrente da incorporação da noção de divulgação científica como parte das responsabilidades de uma ciência cidadã. A terceira perspectiva necessária é o resgate de esforços frente à gestão da informação tecnológica e técnica, ausente das políticas de informação brasileiras a partir dos anos 2000. Associada às perspectivas anteriores, é fundamental que sejam integradas aos esforços de gestão da informação em ambiente digital as práticas tradicionais de bibliotecas e centros de documentação, ou seja, a integração de bases de dados catalográficas e referenciais aos sistemas de gestão da informação digital no contexto da comunicação científica.

Sem a intenção de exaurir possibilidades de gestão integrada da informação científica e tecnológica e o acesso aberto, os autores deste capítulo propuseram um conjunto de linhas de ação que podem beneficiar e integrar, do ponto de vista informacional, as atividades de pesquisa, desenvolvimento e inovação.

## REFERÊNCIAS

- AGUIAR, A. C. Informação e atividades de desenvolvimento científico, tecnológico e industrial: tipologia proposta com base em análise funcional. **Ciência da Informação**, Brasília, v. 20, n. 1, p. 7-15, jan./jun. 1991.
- ALBERTS, B. Engaging in a worldwide transformation: our responsibility as scientists for the provision of global public goods. In: ANNUAL MEETING OF THE NATIONAL ACADEMY OF SCIENCES, 139., 2002, Washington D.C. Disponível em: <<https://www.scienceopen.com/document?vid=b4cae593-1b3f-4e87-abdc-0dc901abb7b7>>. Acesso em: 15 jun. 2017.
- BAILEY, C. W. What is open access? In: JACOBS, N. (Ed.). **Open access**: key strategic, technical and economic aspects. Oxford: Chandos House, 2006. p. 13-26.
- BERLIN Declaration on Open Access to Knowledge in the Sciences and Humanities. Disponível em: <<https://openaccess.mpg.de/Berlin-Declaration>>. Acesso em: 15 jun. 2017.
- BRODY, T. et al. The effect of open access on citation impact. In: NATIONAL POLICIES ON OPEN ACCESS (OA) PROVISION FOR UNIVERSITY RESEARCH OUTPUT: AN INTERNATIONAL MEETING, 2004. **Proceedings...** Disponível em: <<http://opcit.eprints.org/feb19oa/brody-impact.pdf>>. Acesso em: 15 jun. 2017.
- BRODY, T.; HARNAD, S. **The research impact cycle**. 2004. Disponível em: <<http://opcit.eprints.org/feb19oa/harnad-cycle.ppt>>. Acesso em: 15 jun. 2017.
- BUDAPEST Open Access Initiative. 2001. Disponível em: <<http://www.budapestopenaccessinitiative.org/>>. Acesso em: 15 jun. 2017.
- COSTA, M. P.; LEITE, F. C. L. **Repositórios institucionais da América Latina e o acesso aberto à informação científica**. Brasília: Ibict, 2017. Disponível em: <<http://repositorio.unb.br/handle/10482/23202>>. Acesso em: 15 jun. 2017.
- COSTA, S. Filosofia aberta, modelos de negócios e agências de fomento: elementos essenciais a uma discussão sobre o acesso livre à informação científica. **Ciência da Informação**, Brasília, v. 35, n. 2, p. 39-50, maio/ago. 2006. Disponível em: <[http://repositorio.unb.br/bitstream/10482/634/1/ARTIGO\\_FilosofiaAbertaModelosNeg%C3%B3cios.pdf](http://repositorio.unb.br/bitstream/10482/634/1/ARTIGO_FilosofiaAbertaModelosNeg%C3%B3cios.pdf)>. Acesso em: 15 jun. 2017.
- JACOBS, N. (Ed.). **Open access**: key strategic, technical and economic aspects. Oxford: Chandos Publishing, 2006.
- JOHNSON, R. K. Institutional repositories: partnering with faculty to enhance scholarly communication. **D-Lib Magazine**, v. 8, n. 11, nov. 2002. Disponível em: <<http://www.dlib.org/dlib/november02/johnson/11johnson.html>>. Acesso em: 15 jun. 2017.
- LAGOZE, C.; VAN DE SOMPEL, H. The open archives initiative: building a low-barrier interoperability framework. In: ACM/IEEE JOINT CONFERENCE ON DIGITAL LIBRARIES, 1., 2001, Roanoke, Virginia. **Proceedings...** Roanoke, Virginia, 2001. p. 54-62. Disponível em: <<http://dl.acm.org/citation.cfm?doid=379437.379449>>. Acesso em: 15 jun. 2017.
- LAWRENCE, S. Free online availability substantially increases a paper's impact. **Nature web debates**, may, 2001. Disponível em: <<https://www.nature.com/nature/journal/v411/n6837/full/411521a0.html>>. Acesso em: 15 jun. 2017.

LEITE, F. C. L. **Modelo genérico de gestão da informação científica para instituições de pesquisa na perspectiva da comunicação científica e do acesso aberto.** 2011. 262 f. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Brasília, 2011. Disponível em: <<http://repositorio.unb.br/handle/10482/9753>>. Acesso em: 15 jun. 2017.

LEITE, F. C. L.; COSTA, S. M. S. Modelo genérico de gestão da informação científica digital para instituições de pesquisa na perspectiva da comunicação científica e do acesso aberto. **Investigación Bibliotecológica:** Archivonomía, Bibliotecología e Información, v. 30, n. 69, p. 43-74, may/aug. 2016. Disponível em: <[http://ac.els-cdn.com/S0187358X16300168/1-s2.0-S0187358X16300168-main.pdf?\\_tid=22073e7e-3f1a-11e7-b6c7-00000aacb35f&acdnat=1495476893\\_c53e9a6e5f0db5b4720922d3149590cb](http://ac.els-cdn.com/S0187358X16300168/1-s2.0-S0187358X16300168-main.pdf?_tid=22073e7e-3f1a-11e7-b6c7-00000aacb35f&acdnat=1495476893_c53e9a6e5f0db5b4720922d3149590cb)>. Acesso em: 15 jun. 2017.

MIKHAILOV, A. I. On the functioning of the All-Union Institute for Scientific and Technical Information of the USSR Academy of Sciences. In: THE INTERNATIONAL CONFERENCE ON SCIENTIFIC INFORMATION, 1958, Washington, D.C. **Proceedings...** Washington, D.C.: National Academy of Sciences, 1959. v.1, p. 511-522. Disponível em: <<https://www.nap.edu/read/10866/chapter/35>>. Acesso em: 15 jun. 2017.

MORRISON, H. **Scholarly communication for librarians.** Oxford: Chandos Publishing, 2009.

ODLYZKO, A. Economic costs of toll access. In: JACOBS, N. (Ed.). **Open access:** key strategic, technical and economic aspects. Oxford: Chandos Publishing, 2006. p. 39-44.

SUBER, P. **Open access overview:** focusing on open access to peer-reviewed research articles and their preprints. 2010. Disponível em: <<https://legacy.earlham.edu/~peters/fos/overview.htm>>. Acesso em: 15 jun. 2017.

SUBER, P. **Timeline of the Open Access Movement.** 2007. Disponível em: <<http://www.earlham.edu/~peters/fos/timeline.htm>>. Acesso em: 15 jun. 2017.

SWAN, A. **The open access citation advantage:** studies and results to date. Technical Report. Southampton: University of Southampton, 2010. 17 p. Disponível em: <<http://eprints.soton.ac.uk/id/eprint/268516>>. Acesso em: 15 jun. 2017.

WEITZEL, S. R. O papel dos repositórios institucionais e temáticos na estrutura da produção científica. **Em Questão**, v. 12, n. 1, p. 51-71, 2006.

WILLINSKY, J. **The access principle:** the case for open access to research and scholarship. Massachusetts: MIT Press, 2006.







# TECNOLOGIAS PARA GESTÃO DA INFORMAÇÃO

Milton Shintaku





## INTRODUÇÃO

A relação do ser humano com a informação remonta a seus primórdios e muitas vezes ligado a comunicação, à medida que o homem possui as formas mais eficazes de interação. A fala, que é instrumento único do ser humano, a escrita e os símbolos estão repletos de informação, muitas vezes utilizadas para transferir conhecimentos. Assim, essa interação repleta de conhecimento torna-se vital aos seres humanos, que possuem necessidades sociais.

Ainda que não possua unanimidade na definição, o termo informação possui características quase primitivas, ou seja, todos sabem o que é, mas não sabem explicar. Morfologicamente, informação é formada pelo prefixo latino **in** para dentro, **forma** (molde, formato) e **ção** sufixo que apresenta resultado ou processo. Com isso informação é o processo ou resultado de se colocar em uma forma.

Essa definição, de certa forma simplista, alinha-se ao que Oleto (2006) discute ao analisar Buckland (1991), exemplificando a informação como processo, conhecimento e coisa. Assim, informação muda o estado de uma pessoa, reduz as incertezas e faz referência a itens ou adventos existentes. No entanto, a informação é social, que ganha seu real significado em uma comunidade, entre seus elementos.

Ao discutir os estoques e demandas de informação, seus produtos e serviços, revela a necessidade do processo de gestão da informação. Para Barreto (1999) a gestão da informação é a harmonização estética dos estoques de informação existentes e disponíveis, de forma que ocorra a sua assimilação e conhecimento, com técnicas que agregam valor ao todo.

Nesse sentido, numa visão mais organizacional, a gestão da informação deve se apoiar em sistemas de informação adequados às necessidades dos usuários. Com isso, coloca as ferramentas informatizadas como elementos de aplicação das técnicas da gestão da informação.

Essa visão mais técnica é complementada por Marchiori (2002), que amplia a visão da gestão da informação, como área interdisciplinar na qual abrange vários processos, interagindo com outras áreas da ciência da informação. Assim, a gestão da informação tem relação íntima com a tecnologia e suas ferramentas, mas interage com várias linhas e disciplinas como a biblioteconomia, arquivologia, linguística e comunicação.

Nesse contexto, este capítulo trata de tecnologias voltadas à gestão da informação de forma mais ampla, de forma a descrever softwares que apoiam diversas atividades relacionadas à organização, representação, disseminação e outros. Com isso, apresenta opções tecnológicas, de forma a apresentar o cenário complexo que envolve as ações relacionadas à gestão da informação.

## TECNOLOGIAS PARA A GESTÃO DA INFORMAÇÃO

A gestão da informação é definida pelo *Manual de Gestão da Informação*, publicado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) e pelo Instituto de tecnologia do Paraná (Tecpar), como a atividade de gestão e pode ser considerada como um conjunto de processos que englobam atividades de planejamento, organização, direção, distribuição e controle de recursos de qualquer natureza, visando à racionalização e à efetividade de determinado sistema, produto ou serviço. Com isso, envolve diversas atividades que permeiam outras atividades (MANUAL..., 1997).

### SISTEMA INTEGRADO DE GESTÃO DA BIBLIOTECA KOHA

No caso da biblioteca, nota-se certo alinhamento entre alguns processos vinculados à gestão da informação e algumas atividades automatizadas pelos Sistemas Integrados de Gestão de Biblioteca (SIGB). Com isso, as ferramentas de SIGBs tornam-se importantes para as atividades de gestão da informação, visto que tornam-se meios para aplicar conceitos.

Nesse sentido, o software livre de SIGB mais utilizado no mundo é o Koha. Conforme o seu portal, existem 976 bibliotecas utilizando essa ferramenta por todo o mundo, como mostra a Figura 1. Entretanto, essa ferramenta ainda não é tão conhecida no Brasil, com poucos usuários, sendo que na América do Sul a Colômbia apresenta o maior quantitativo de bibliotecas usuárias.

A Índia é um grande usuário do Koha, colocando o Sul da Ásia como maior área de uso, seguido da Europa e Oceania. Na Europa, a Alemanha e Reino Unido formam o grupo de maiores usuários. A Oceania justifica-se por ser o berço do Koha, visto que foi desenvolvido pela Biblioteca Horowhenua da Nova Zelândia, lançado em 2000, tornando-se mundial pelas contribuições da comunidade de profissionais e usuários de todo o mundo que participam ativamente de seu aperfeiçoamento, denominado de Koha Community. Yang e Hoffmann (2010) comparam alguns SIGBs e revelam que o Koha atende a maior parte das necessidades das bibliotecas, no que diz respeito à gestão de catálogos online.

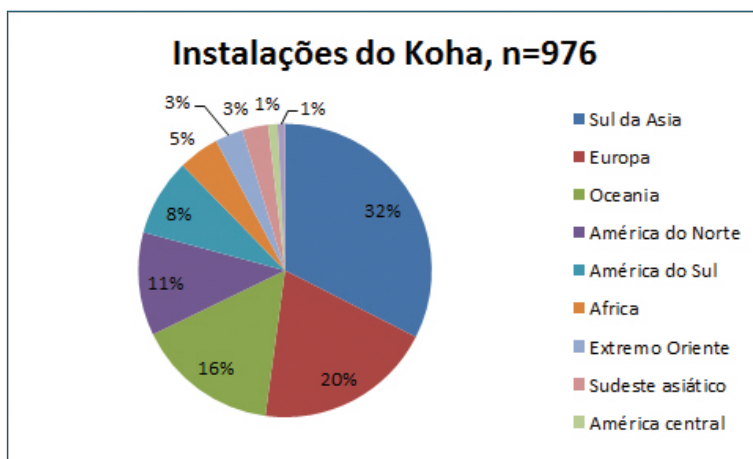


Figura 1 – Uso do Koha no mundo

Fonte: Koha Wiki (2015) .

Como mostra a Figura 2, o Koha possui módulos que atendem a grande parte das atividades de uma biblioteca, abrangendo a gestão do catálogo e atividades relacionadas, como aquisição, catalogação, circulação e outros. Com isso, torna-se uma opção viável para ser implementado em bibliotecas que possuem certa restrição orçamentária, que dificultam a aquisição de licenças de softwares proprietários.

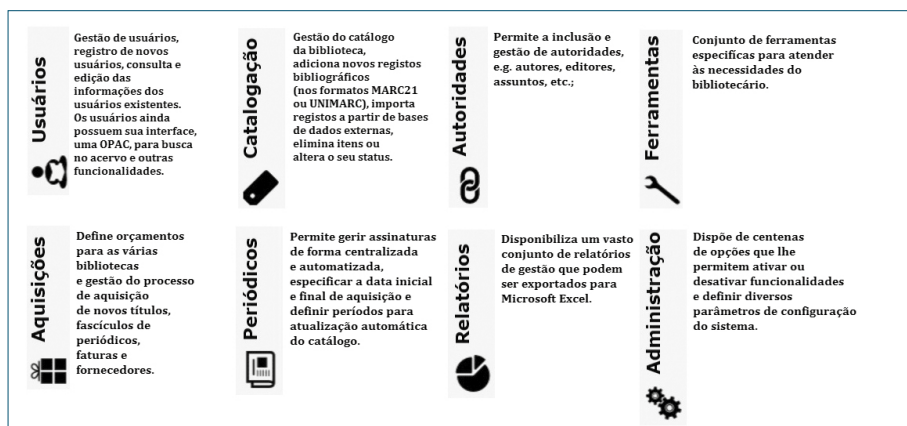


Figura 2 – Módulos do Koha

Fonte: Schiessl et al. (2016).

O Koha pode ser utilizado por instituições que possuem mais que uma biblioteca, em uma única instalação do software, mantendo a independência das informações mantidas pelo Koha. Schiessl et al. (2016) em seu trabalho apre-

senta o Koha operando totalmente na web, possibilitando a gestão distribuída com a opção de múltiplas bibliotecas. Destacam-se, também, as possibilidades de interação entre o Koha e outros sistemas de informação. O Koha implementa o protocolo Z39-50, permitindo a oferta de serviços como a catalogação colaborativa. Da mesma forma, responde ao protocolo *Open Archive Initiative - Protocol Metadada Harvesting (OAI-PMH)*, permitindo a interoperabilidade por coleta automática de metadados.

Por ser um software livre, o Koha apresenta-se apropriado para bibliotecas públicas ou escolares. Egundjobi e Awoyemi (2012), em análise do uso na Nigéria, demonstram a satisfação com o software por parte da equipe de bibliotecários. Anuradha, Sivakaminathan e Arun Kumar (2011), na visão de bibliotecas indianas, evidenciam que o Koha pode integrar-se a outros softwares livres. Amante e Marçal (2012), numa experiência de Portugal, utilizam o Koha em bibliotecas universitárias.

Atualmente no Brasil, o Koha é a opção em software livre mais viável. Nesse sentido, o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), juntamente com a Secretaria Nacional de Juventude (SNJ) desenvolveram projeto para a sua implementação e disseminação a órgãos do governo, que necessitem de SIGB. Em suma, o Koha se apresenta como a opção mais atual para SIGBs em software livre, oferecendo serviços para a gestão de acervos, além de possibilidades de interoperabilidade com outros sistemas. Operando totalmente pela internet, o Koha possibilita a gestão distribuída de múltiplas bibliotecas e, com isso, apropriadas a vários tipos de bibliotecas, principalmente as públicas e escolares, que possuem certas restrições orçamentárias.

## SISTEMAS DE PUBLICAÇÃO

Uma parte importante na gestão da informação reside em sua interface com a comunicação da informação, principalmente nas atividades voltadas às publicações. No caso das publicações científicas, o modelo de comunicação, como o desenvolvido por Bjork (2007), apresenta canais diferentes para a publicação. Para essa atividade, como mostra a Figura 3, deve-se ter ferramentas diferentes, à medida que possuem publicações diferentes.

Um manuscrito pode conter vários formatos, como livros, artigos, resumos expandidos, cada qual possui características diferentes. É de conhecimento que algumas disciplinas possuem preferência por certos canais, as ciências humanas, por exemplo, possuem predileção pelos livros. Cientistas da computação gostam de eventos como congressos ou conferências, que possuem maior rapidez em

todo o processo. Grande parte das disciplinas utilizam o artigo de periódico para publicar seus resultados de pesquisa.

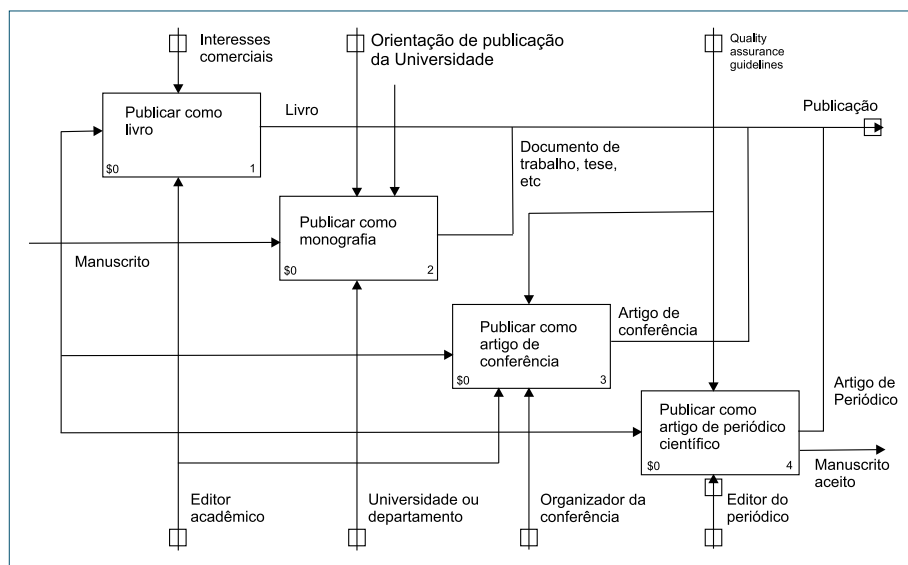


Figura 3 – Diagrama de publicação de resultados de pesquisa

Fonte: Bjork (2007).

Nesse sentido, o *Public Knowledge Project* (PKP) desenvolve uma linha de ferramentas livres voltadas à publicação científica, alinhado ao movimento de acesso aberto (*open access*). Com isso, contempla todos os principais canais de disseminação, a publicação de livros, periódicos científicos e artigos de eventos, atendendo os canais preferenciais de todas as disciplinas.

### Ferramenta para Publicação de Livros: *Open Monograph Press*

Independentemente do processo de digitalização, o livro ainda é um grande canal de disseminação da informação. Em alguns casos fica quase impossível separar entre livros técnicos, científicos ou de literatura. Nos casos de história, por exemplo, um livro pode ser resultado de pesquisa (logo, científico), mas que tenha apelo literário, ou seja, utilizado como material didático. Assim, livros físicos em papel têm convivido com os livros digitais (*e-books*).

Para o processo editorial do livro digital, o PKP desenvolveu o *Open Monograph Press* (OMP). Esta ferramenta cria o portal da editora e automatiza o fluxo editorial do livro, facilitando a sua publicação digital. Assim, atende-se a insti-



tuições que queiram disseminar livros digitais ou para editoras informatizarem seus processos.

O OMP é mais flexível, à medida que possibilita às editoras aportar novos atores no processo de publicação, dando maior vazão a materiais que estão fora do mercado comercial, de acordo com Willinsky (2009). Esse ponto contribui com o cenário atual, no qual as publicações digitais apresentam maior diversidade de tipos e formatos.

O OMP não opera somente com livros em acesso aberto, como descreve Adema (2010), em análise do uso de *e-books* nas humanidades e ciências sociais.

Adema (2010) relata que o OMP, mesmo que fundamentado em acesso aberto, possui funcionalidades que possibilitam um modelo econômico híbrido, no qual se comercializem algumas publicações. Essa flexibilidade torna-se útil para muitas editoras. No Quadro 1, destaca-se como o OMP atua em várias atividades de uma editora, facilitando a sua operacionalização, na medida em que possui módulos que atendem a atividades distintas, para usuários diferentes.

Quanto às vantagens na utilização do OMP, Thiagarajan (2013) enumera que a ferramenta é robusta, sem custos de licença, possibilita a interoperabilidade, possui fluxo de publicação simples e pode ser instalado em vários sistemas operacionais. Adiciona-se a essas vantagens a questão de ser uma ferramenta padronizada internacionalmente.

O OMP foi adotado pela Universidade de São Paulo (USP), no seu Portal de Livros Abertos<sup>1</sup>. O Diretor da Divisão de Gestão de Sistemas de Comunicação e Disseminação de Produtos e Serviços do Sistema Integrado de Bibliotecas da USP, André Serradas, relata que a escolha do OMP se deu em primeiro lugar por ser um sistema específico para a gestão de livros e com as ferramentas e metadados específicos, como o *ONline Information eXchange* (ONIX). Entretanto, o que pesou bastante foi o fato de ser uma iniciativa do PKP, muito diferenciada em relação à dedicação aos projetos

De forma geral, o OMP se apresenta como uma ferramenta viável para automatizar as atividades de uma editora, que publica livros alinhados ou não ao acesso aberto, apresentando mais vantagens e flexibilidade ideais que vários tipos de instituições. O OMP, também, oferece a interoperabilidade, possibilitando que a instituição participe de redes de editoras, apropriado a integração de informações automática de sistemas.

---

1 Disponível em: <<http://www.livrosabertos.sibi.usp.br/portaldelivrosUSP>>. Acesso em: 23 jul. 2017.



MÓDULO	USOS	COMPONENTES	FUNÇÕES
CARGA	Manuscrito Planos de Marketing Contratos de autor	Metadados Arquivos Controle de versão	Oferta funções voltadas ao carregamento e ao gerenciamento de vários tipos de arquivos, com customização de metadados, instruções e direitos de acesso.
REVISÃO	Revisores Internos Revisores externos Múltiplas rodadas	Seleção de revisores Formulário de revisão Preparação de e-mails	Renomear processos, criar formulários de revisão, ajustar e-mails, estabelecer processo de revisão mais interativo.
ASSINATURA	Editor/autor Editor/marketing Edição/autor	Preparar e-mails Designar usuários Designação	Assegurar notificações e papéis.
PRODUÇÃO	Edição Design Marcação	Naming options File management Web publication	Permite operar em um fluxo com arquivos em vários formatos, como requerido pelo processo de publicação (original, XML, PDF, doc.).
USUÁRIOS	Corpo editorial Autores e leitores Revisores	Designação Tabela de itens Enroll/ un-enroll	Informação dos usuários e seus papéis no sistema.
TRANSAÇÃO	Assinatura Compra Acesso	Cartão de crédito Lista de membros Tabela de preços	Permite a usuários pedir e comprar itens, com opção de filiação.
EDITORA	Site da Editora Página do livro Catálogo de livro	Publicações Notícias Busca	Portal de gestão de publicações.

Quadro 1 – Funcionalidades do OMP

Fonte: Brito e Shintaku (2016).

### Ferramenta para Publicação de Anais de Eventos: *Open Conference System*

Os eventos científicos são elementos importantes na disseminação da informação, muito por seu caráter mais dinâmico e prático. Para autores como Lievrouw (1990), os eventos atuam de forma intermediária, com abrangência mais restrita e não necessariamente apresentando os resultados finais de pesquisa. Bjork (2007), por sua vez, considera os eventos como comunicações finais, assim como as publicações em periódicos científicos.

Nesse sentido, o PKP desenvolveu o *Open Conference System (OCS)*, no Brasil também conhecido como Sistema de Organização e Acompanhamento de Conferências (SOAC), para criar portais de eventos, que automatizam algumas atividades. Com isso, atende as necessidades de instituições que organizam eventos científicos, assim como a disseminação dos trabalhos apresentados.

O OCS é uma ferramenta com grande potencial para gestão, avaliação e difusão de trabalhos acadêmicos, destacando-se pela sua evolução que incorpora novas funcionalidades. Relatam Maquilón Sanches, Lillo Hidalgo e Mirete Ruiz (2011) que o OCS torna-se uma ferramenta de apoio às instituições de ensino e pesquisa, na disseminação das informações.

O OCS possibilita criar portais de eventos nas instituições, como apresentado na Figura 4. Em uma mesma instalação do OCS pode-se criar sites de vários eventos. Em cada site de evento terá as suas ocorrências, assim como, para cada ocorrência do evento terá as suas páginas ofertando serviços. Assim, com uma única instalação pode resolver os problemas para os sites de eventos na instituição.

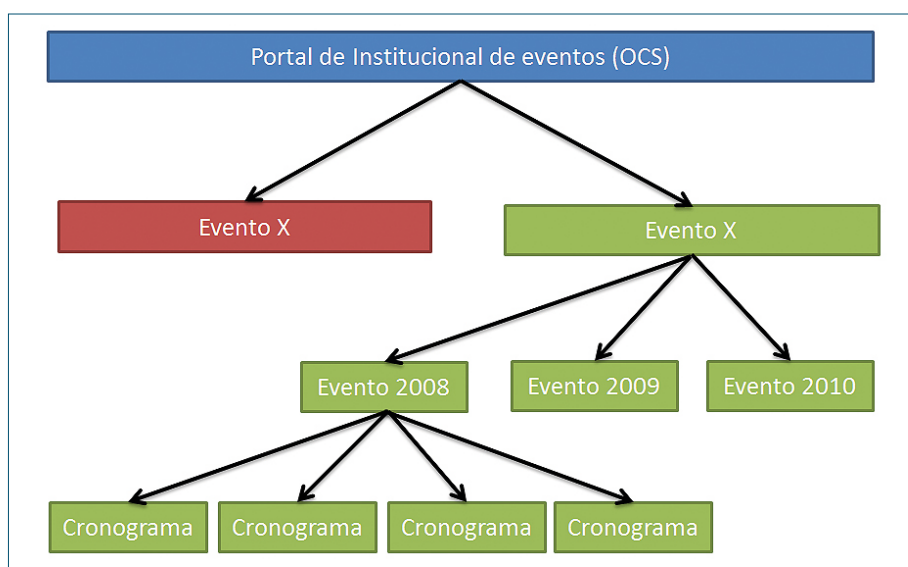


Figura 4 – Diagrama de portal de eventos com o OCS

Fonte: Shintaku, Brito e Fleury (2015).

Para possibilitar a operacionalização de um portal de eventos, mantendo o particionamento entre os diversos eventos hospedados, o OCS possui uma forma organizada e hierárquica de papéis, como apresentados por Shintaku, Brito e Fleury (2014), no Quadro 2. Cada papel possui suas permissões e atua em determinada atividade no portal, evento, ocorrência do evento ou serviço.

O OCS torna-se uma opção prática para instituições que precisam gerenciar eventos científicos, por meio de sites, com a oferta de serviços como submissão de trabalhos, avaliação e publicação de anais, ideais para as universidades e institutos de pesquisas. Simplifica as tarefas da equipe de informática, à medida que utiliza apenas uma instalação, mas com a possibilidade de inúmeros sites, com independência de configurações e identidades visuais.

PAPEL	ATRIBUIÇÕES
Administrador	Configurar o portal de eventos Criar conferências; Designar gerentes gerais
Gerente geral	Criar edições de uma conferência; Definir o cronograma; Elaborar formulários de avaliação personalizados; Habilitar ferramentas de leitura
Diretor	Conduzir a avaliação por pares; Notificar autor; Elaborar anais;
Diretor de modalidade	Conduzir a avaliação por pares em uma modalidade.
Avaliador	Listar trabalhos atribuídos para avaliação; Preencher formulário de avaliação.
Autor	Submeter trabalho; Enviar correções.
Participante	Pagar inscrição; Acessar o programa da conferência; Ver informações sobre hospedagem e localização do evento.
Leitor	Acessar os trabalhos publicados nas conferências. Não possui nenhum privilégio no sistema.

Quadro 2 - Papéis e atribuições

Fonte: Shintaku, Brito e Fleury (2014).

### Ferramenta para Publicação de Periódicos Científicos: *Open Journal System*

Os artigos de periódicos científicos são considerados como canais naturais para a disseminação da informação, tanto que, no modelo de Garvey e Griffith (1972), são apresentados como o meio pelo qual o conhecimento científico é apresentado no formato final. Desde o seu surgimento no século XVII, os periódicos científicos mantêm o status de canal preferencial de publicação de muitas disciplinas.

Assim, o PKP desenvolveu o *Open Journal System* (OJS) para atender a criação e portais de periódicos científicos, contemplando todas as estruturas e serviços, oferecendo facilidades para a gestão das revistas, principalmente para as tarefas do editor. Com isso, em uma única instalação pode-se criar várias revistas científicas, com suas características próprias, num modelo mais econômico e eficaz.

O OJS oferta várias funcionalidades voltadas à publicação de revistas, tanto que Bergamaschi e Duarte (2012) destacam as seguintes: submissão de artigos e pareceres; gerenciamento online de cada fase da publicação; indexação com-

pleta dos artigos publicados; ferramenta de ajuda para leitura e pesquisa dos artigos. Torna-se, assim, uma ferramenta completa para a criação de periódicos científicos, principalmente de acesso aberto.

O OJS não objetiva apenas à criação de periódicos, mas também à migração de revistas já existentes. Carlos (2009) destaca a possibilidade de publicação dos fascículos anteriores, possibilitando a publicação de todas as edições da revista, da mesma forma que possibilita disponibilizar online, inclusive, revistas já descontinuidas, com suas edições preservadas.

Bergamaschi e Duarte (2012) afirmam que o OJS é um modelo alternativo de ampliar o acesso, visibilidade, impacto e preservação dos resultados publicados nos artigos. Com isso, a linha a que Baptista et al. (2007) advoga, afirmando que o OJS é a principal ferramenta para a criação de revistas de acesso aberto. Essas afirmações encontram reforços nos resultados obtidos por Shintaku, Brito e Carvalho Neto (2014), em relação à quantidade de revistas brasileiras desenvolvidas com o OJS e suas indexações.

Pires e Klebersson (2012) atribui a adesão das revistas ao OJS a: redução dos custos de publicação; facilidade de acesso e consulta via web; maior visibilidade aos artigos; crescimento dos grupos de pesquisa e preocupação dos programas de pós-graduação em publicar suas revistas; incentivo e apoio financeiro às revistas por órgãos de fomento. Miranda, Moraes e Barros (2010) afirmam que o SEER/OJS ajuda na qualidade das revistas, ao passo que padroniza as revistas hospedadas em um portal, dando transparência aos processos editoriais.

Diante de tantos pontos positivos, o OJS torna-se a ferramenta mais apropriada para a publicação de periódicos científicos de acesso aberto no contexto brasileiro. Conta-se, também, com o apoio oferecido pelo Ibict, parceiro do PKP, que transfere a tecnologia, por meio de treinamentos e documentação técnica.

## REPOSITÓRIOS

No contexto do movimento de acesso aberto, duas iniciativas se destacam, os sistemas de publicação (representado pelos periódicos científicos) e os repositórios, no que ficaram denominados por Harnad et al. (2004) como via dourada e verde. Assim, as novas informações científicas são publicadas nos periódicos, mas podem ser preservadas e acessadas nos repositórios.

Esse conceito inicial justifica o nome de repositório, que, morfológicamente, significa, **re** (novamente) + **por** (colocar) + **tório** (local), ou seja, local em que se repõe, ou se recolocam coisas. Isso, porque nos repositórios haveria cópias de publicações já feitas. Entretanto, conforme Shintaku e Vidotti (2016), esse contexto mudou, colocando os repositórios como primeira fonte, como nos casos das teses e dissertações.

Outra mudança considerável é a expansão dos conceitos de repositórios para outras disciplinas relacionadas à informação, como a arquivologia e museologia, pela necessidade de ferramentas para gestão da informação digital. Com isso, mais ferramentas foram incorporadas às existentes, de forma a abranger todas as formas de informação e apoiar diversas atividades.

### Repositórios para Bibliotecas: *DSpace*

Os repositórios para bibliotecas remontam ao final do século passado, ainda com a denominação de bibliotecas digitais. Entretanto, os repositórios receberam destaque maior com o movimento de arquivos abertos, acesso aberto e outros, principalmente pelas possibilidades de interoperabilidade e autoarquivamento, o qual permitiu troca de informações entre sistemas e maior participação do autor na disseminação da informação.

Atualmente, a ferramenta mais utilizada mundialmente para construção de repositórios é o *DSpace*, desenvolvida pela Hewlett-Packard (HP) para a bibliotecas do *Massachusetts Institute of Technology* (MIT). Posteriormente, tornou-se um software livre, apoiado por uma grande comunidade, e mantido pela *DuraSpace*. No Brasil, o Ibict apoia o *DSpace* com treinamentos, tradução e produção de documentação técnica.

Lynch (2003) afirma que o modelo gerado no MIT, com o *DSpace*, é apropriado para ser utilizado por outras instituições, na medida em que é um software livre, sem barreiras, de código livre. Nota-se que as necessidades de disseminação da informação das bibliotecas são semelhantes nas instituições de ensino e pesquisa, mesmo com as diferenças em termos de país, organização e outros.

Tansley et al. (2003), apresentando o *DSpace* e suas funcionalidades, relata que o interesse pelos repositórios por parte da comunidade fomenta a confiança no desenvolvimento de funcionalidades voltadas à preservação e gestão de conteúdos digitais. Com isso, garante-se a evolução e sustentabilidade do *DSpace*, como um software livre.

Na questão das funcionalidades, o Quadro 3 apresenta algumas características do *DSpace* que apoiam o seu uso. No caso, tudo pode ser ajustado conforme as necessidades da instituição na disseminação de conteúdo. Pode-se combinar as características para obter o resultado desejado, adaptando o *DSpace* às suas necessidades.

FUNCIONALIDADE	USO	ITENS	FUNÇÕES
Estrutura informacional	Organização do conteúdo	Comunidade Subcomunidade Coleção Item Arquivos	A organização do DSpace é hierárquica e repetitiva, ou seja, um DSpace pode ter inúmeras comunidades. Estas, por sua vez, podem ter várias subcomunidades, e assim por diante
Alimentação	Entrada de conteúdos	Submissão Importação Coleta automática	O DSpace oferece vários meios pelos quais podem se inserir conteúdos, manuais, automático ou em lote, de forma a facilitar o processo
Fluxo de submissão	Controle da submissão	Depósito Avaliação Correção de metadados	Se a opção for a submissão, pode-se ajustar as etapas (fluxo) pelas quais um conteúdo deve passar até ser disponibilizado
Recuperação	Obtenção de conteúdos	Busca Navegação Exportação Resposta a coleta	Há várias formas de recuperar conteúdos, manuais, automáticas, em lote, por busca e por refinamento
Usuários	Gerenciar os usuários	Administrador Anônimo Usuário	Os usuários no DSpace estão intimamente relacionados aos serviços, permissões e acessos
Permissões	Controlar o acesso às funcionalidades	Leitura Criação Remoção Adição	Os recursos ou conteúdo no DSpace requerem permissões para gestão
Acesso	Controlar o acesso aos conteúdos	Aberto Restrito Embargado	Um conteúdo pode ser de livre acesso, restrito a um grupo de usuários ou embargado por tempo limitado

Quadro 3 – Características *DSpace*

Fonte: Autoria própria (2017).

Possivelmente aliado à oferta de funcionalidades, a aceitação do *DSpace*, deve-se, muito, a sua flexibilidade, à medida que pode ser implementado em vários tipos de instituições, com finalidades diferentes. Com isso, pode ser utilizado em vários tipos de instituições. Murakami e Fausto (2013) verificaram o uso em universidades e institutos de pesquisa; já Macedo, Shintaku e Brito (2015), no âmbito governamental, corroborando essa afirmação.

Assim, conclui-se que o *DSpace* se apresenta como uma opção viável e confiável para bibliotecas disseminarem conteúdos sob sua tutela, independentemente do tipo de instituição, ou seja, não apenas as acadêmicas, em um papel mais ativo. O *DSpace* é um software livre apresentado como caso de sucesso mundial.

### Repositórios para Arquivos: *Archivematica*

A Tecnologia da Informação e Comunicação (TIC) tem atuado na administração, em quase todos os processos. Não raro, equipes de informática são destinadas a manter sistemas voltados à gestão administrativa. Entretanto, no que diz respeito à informação administrativa, não se apresentou o mesmo empenho, muito possivelmente por questões de certificação, entre outros fatores.

Com as mudanças e evolução, torna-se cada vez mais comum que documentos administrativos sejam criados, tramitados e arquivados em formato digital. Dessa forma, tem-se a necessidade de ferramentas que possam atender aos arquivos, conforme as características desse tipo de informação – nesse caso, repositórios para documentos de cunho arquivístico, que respeitem as suas características.

Lampert (2016), analisando o *DSpace*, verificou que não atende as especificidades dos documentos arquivísticos, mesmo que seja uma ferramenta para repositórios. Para o mesmo autor, o *Archivematica* se apresenta como opção válida para a criação de repositórios arquivísticos, por atender melhor as questões de preservação, primordial aos arquivos, mesmo que não atenda totalmente as questões de acesso.

O *Archivematica* agrega serviços de curadoria e preservação. Van Garderen (2010) descreve a arquitetura e serviços oferecidos pelo *Archivematica*, apresentado na Figura 5. Destaca-se, dessa forma, que, desde a entrada de um documento até o seu armazenamento final, passa-se por vários processos, de forma a garantir a curadoria e preservação.

Conforme Van Garderen (2010), pode-se agrupar os serviços ofertados pelo *Archivematica* em três grupos: submissão, arquivamento e disseminação. Assim, as tarefas de receber, revisar, quarentena e apreciar são ligadas à submissão, por tratar da entrada do documento. Já as tarefas de preparar, revisar e armazenar ligam-se ao arquivamento, enquanto prover e monitorar fazem parte da etapa de disseminação.

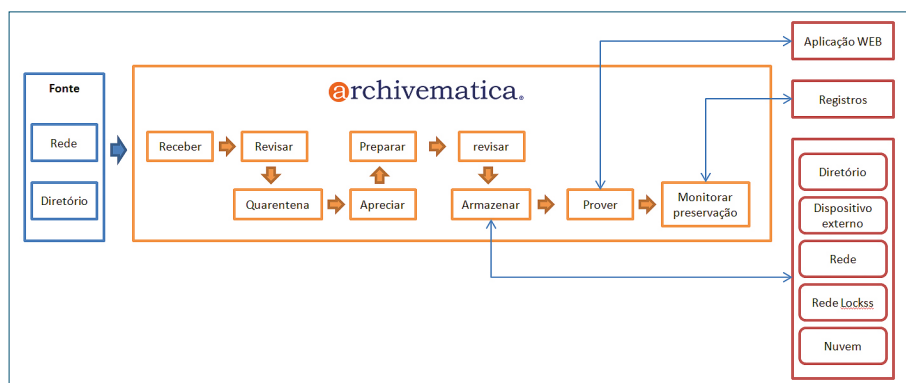


Figura 5 – Arquitetura do Archivemática

Fonte: Baseado em Van Garderen (2010).

O processo de submissão envolve uma série de tarefas, como as questões de descrição dos metadados, tendo a possibilidade do uso do *Dublin Core*, descompactação, checagem, identificação, identificação automática de formato, validação do formato, normalização, quarentena, entre outros. Essas tarefas asseguram que os documentos entrados tenham um tratamento apropriado para as próximas etapas.

Estratégias de preservação devem manter as características dos documentos. Assim, para cada formato de documento, deve-se ter formas diferenciadas para armazenar, da mesma forma que o armazenamento pode ser efetuado de diferentes maneiras, de diretórios de arquivos até nas nuvens, passando por técnicas como de múltiplas cópias do *Lot of Copies Keep Stuff Safe* (LOCKSS).

A disseminação do Archivemática, como relata Lampert (2016), não oferece tantas facilidades quanto outras ferramentas de repositórios. Por isso, associa-se ao Archivemática outra ferramenta que apoie os processos de recuperação, visto que o forte dessa ferramenta é a curadoria e preservação. No presente caso, optou-se pelo *Access To Memory* (AToM), que se integra bem ao Archivemática.

O AToM também é uma ferramenta livre, software aberto, desenvolvido conforme as normas internacionais arquivísticas. Van Garderen (2010) descreve o AToM como um software baseado em web, que aproveita as facilidades oferecidas por essa tecnologia, permitindo ao arquivista ofertar novos serviços aos usuários. Jorente (2016) revela que o AToM e Archivemática se integram de forma complementar, fazendo curadoria, preservação, representação e disseminação.

Assim, para atender as atividades primordiais de arquivos, relacionados à preservação, o Archivemática se torna apropriado. Entretanto, para o momento tecnológico atual, o AToM se faz necessário, visto que leis e normas exigem que documentos de cunho arquivístico estejam disponíveis na internet. Com isso, o consórcio Archivemática - AToM torna-se a opção mais apropriada.



## Repositórios para Museus: *Omeka*

As atividades relacionadas à informação da museologia compartilham muitas semelhanças com as da biblioteca e arquivo. Fazer o tratamento e curadoria de itens de um acervo lembra em parte tratamento de catálogos de obras raras ou mapas históricos na biblioteca. Como grande parte dos itens de museus são únicos, assemelham-se a documentos arquivísticos. Entretanto, museus são únicos e cada vez mais requerem ferramentas específicas que apoiem suas atividades.

Para a criação de repositórios de cunho museológico, Martins et al. (2015), discutindo os aspectos amplos da finalidade dos museus acadêmicos, avaliaram Sistemas Integrados de Gestão de Museus (SIGM), definindo o *Omeka* como opção apropriada, principalmente pela possibilidade de integração e facilidade do uso. Mesmo que não ofereça todas as funcionalidades necessárias, o *Omeka* apresenta-se operacional para a gestão de acervos museológicos.

Da mesma forma, Kucsma, Reiss e Sidman (2010) testaram o *Omeka* para disseminação de coleções digitais e chegaram à conclusão de que a ferramenta, apesar de algumas limitações, era apropriada para atender a pequenas e médias instituições na disseminação de conteúdos digitais, como em museus web. Entretanto, o autor revela que o *Omeka* se apresenta promissor, visto que a sua evolução pode resolver os problemas que possam aparecer.

Scheinfeldt (2008) relata que as funcionalidades ofertadas pelo *Omeka* apoiam as exposições online, no qual os profissionais de cultura podem ter papel mais ativo. Destaca-se que o *Omeka* possui curva de aprendizado curto, ou seja, é de fácil utilização, com funcionalidades voltadas para coleções de patrimônio cultural, sendo, com isso, apropriado a museus.

Nesse sentido, o *Omeka* se apresenta com maiores funcionalidades que os repositórios para bibliotecas e arquivos, com maior expressividade na apresentação dos conteúdos, principalmente para imagens. Cohen (2008) destaca que o *Omeka* possui uma interface favorável a exposição e um núcleo robusto para criação de sites de disseminação de conteúdo.

Pérez (2011) menciona a importância de exposições digitais, pela sua capacidade de comunicação, e o *Omeka* apoia esse processo de forma fácil. Assim, cria exposições estruturadas online, com descrição dos itens apresentados de forma econômica, com grande abrangência. No caso de museus, pode-se criar coleções digitais sobre os itens físicos e expô-las no *Omeka*, de forma a atrair novos visitantes, que queiram ver os originais.

Portanto, o *Omeka* se apresenta como uma plataforma de publicação de conteúdo na web, apropriada a exposições, adequada a museus. Assim, pode

apresentar coleções no formato digital de forma gráfica. Possivelmente, a grande vantagem do *Omeka* é a facilidade de uso e as possibilidades dadas aos usuários na publicação de conteúdo.

### Repositórios para Dados: *Comprehensive Knowledge Archive Network*

Repositórios têm despertado o interesse das instituições por diversos motivos. Entretanto, mesmo que mantenham conceitos semelhantes, repositórios ganham especificidade dependendo da sua finalidade. Como visto nas seções anteriores, pode-se ter uma ferramenta distinta para cada finalidade, ou seja, repositórios específicos para tipos de informação, ferramentas distintas para cada tipo de repositório.

Nesse sentido, os dados têm ganhado destaque, à medida que leis e movimentos clamam por sua disseminação. Para a ciência, o Movimento dos Dados Abertos (*Open Data*) advoga a liberação e disseminação dos dados de pesquisa, não apenas os resultados. No governo brasileiro, os dados fazem parte da Lei de Acesso à Informação.

Assim, uma opção de ferramenta para criação de repositórios de dados é o *Comprehensive Knowledge Archive Network* (CKAN), desenvolvido e mantido pela *Open Knowledge Foundation*. O CKAN tem por objetivo armazenar coleções organizadas de dados, facilitando a recuperação e apresentação. Com isso, possibilita a gestão de bases de dados, apropriadas a instituições que desejam criar repositórios de dados.

O CKAN é uma ferramenta livre, que como todos, oferta a infraestrutura necessária a publicação de bases de dados. Esse ponto encontra alinhamento com Corrêa, Corrêa e Silva (2014) que advogam pelo uso do CKAN para publicação e gestão de dados governamentais abertos, num contexto evolutivo, na medida em que muda uma cultura, alterando procedimentos tradicionais em relação aos dados.

O movimento de dados abertos implica em uma mudança de atitude reque-rendo softwares que o atendam. Son et al. (2014) descrevem o CKAN como uma plataforma operando na internet para disponibilizar dados a todos, numa forma de catálogo. A Figura 6, Son et al. (2014), representa o fluxo dos dados no CKAN: os usuários depositam e recuperam as bases de dados por meio de interface web, organizados em um catálogo formado pelas bases de dados, metadados, registros históricos, dados geoespaciais, tudo validado por um processo interno.

O CKAN organiza e disponibiliza conjuntos de dados (*datasets*) apoiado por metadados (descrições) em vários formatos. Um conjunto de dados pode ter mais que um arquivo (chamado de recurso), mesmo que tenham formatos diferentes.

Assim, a entrada no catálogo é do conjunto de dados, que pode conter vários recursos em formatos distintos. Esses recursos podem ser visualizados diretamente no CKAN.

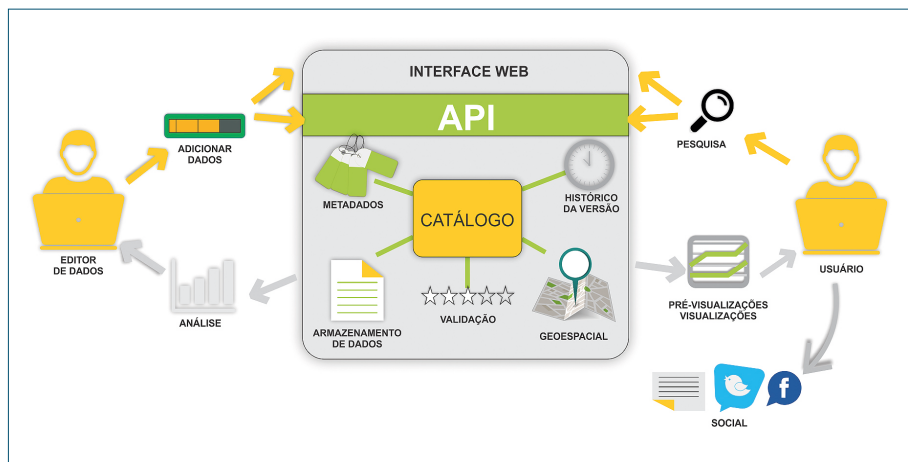


Figura 6 – Fluxo dos dados no CKAN

Fonte: Baseado em Son et al. (2014).

Cabe ressaltar que o CKAN foi a ferramenta adotada pelo projeto de Dados Abertos Governamentais do Ministério do Planejamento, Orçamento e Gestão (MPOG). Com isso, torna-se a opção mais viável à criação de repositórios de dados, principalmente os governamentais com bases de dados estatísticos, organizado em forma de catálogo de conjuntos de dados.

## SISTEMAS DE DESCOBERTA E ENTREGA: VUFIND

Outro ponto importante na gestão da informação é a criação de ecossistemas informacionais com a integração de sistemas independentes. Para tanto, uma das formas de execução é o uso de sistemas de descoberta e entrega, que Vaughan (2011) descreve como uma ferramenta capaz de conectar usuários a grande quantidade de informação provenientes de várias fontes.

Nesse mesmo sentido, Sutradhar (2014) menciona que os sistemas de descoberta e entrega são tecnologias transformativas, possibilitando aos usuários acesso a uma vasta gama de conteúdos coletados e indexados, de forma ranqueada, numa interface com usabilidade. Alinhado a esses preceitos, o *VuFind*, desenvolvido e mantido pela Universidade de Villanova (Pensilvânia, Estados Unidos), revela-se como um sistema de descoberta e entrega em software livre.

Emmanuel (2011) descreve o *VuFind* como um avanço nos serviços ofertados pelas bibliotecas, como em um catálogo online modernizado. Com isso, contextualiza os sistemas de descoberta e entrega aos serviços de bibliotecas, na intermediação da informação. Denton e Coysh (2011) revelam que a interface do *VuFind* possui usabilidade, no qual os usuários navegam intuitivamente, melhorando os serviços de biblioteca.

Shintaku e Macedo (2015), descrevendo as funcionalidades do *VuFind*, destacam:

- a) coleta: utiliza diversas formas de coletar metadados dos sistemas, tanto nos padrões de metadados (*marc21*, *marcxml*, *Dublin Core*), quanto nos protocolos (OAI-PMH, Z39.50), integração com diversos catálogos e integração com outros serviços de descoberta (EBSCO, Primo, Summon);
- b) indexação: utiliza ferramentas como SOLR para indexar os metadados coletados, com aspectos mais técnicos, em que todas as palavras se tornam ponto de recuperação. Podem-se criar índices exclusivos para alguns campos, como título e autor;
- c) recuperação: a busca se dá tanto por ferramenta de busca simples como avançada, além da navegação por refinamento das facetas, podendo-se realizar várias configurações na busca e facetas;
- d) apresentação: adota a apresentação próxima a catálogos de bibliotecas, mas com possibilidades de integração com ferramentas externas, como o *Google books*, que permite visualizar parte do conteúdo, ou mesmo a capa do livro. São apresentadas sugestões de recomendações, bem como documentos relacionados;
- e) interação: permite interação social com inserção de comentários, favoritos e *tags*; personalização dos formatos de citações para os documentos; compatibilidade com Zotero e outras funcionalidades.

Assim, para formar o ecossistema informacional o *VuFind* utiliza protocolos de comunicação como o OAI-PMH (Figura 7). Com isso, pode-se coletar informações de todos os tipos de sistemas que implementam esse protocolo. Uma interface web de busca apoia a descoberta das informações coletadas.

O *VuFind* revela-se um software promissor, à medida que as instituições precisam integrar os seus sistemas de informação, de forma a obter resultados de busca mais consistentes. Com a evolução dessa ferramenta, pode-se esperar que contemple outras funcionalidades, principalmente, relacionadas à gestão da informação consolidada, visto que é uma necessidade atual dos gestores.

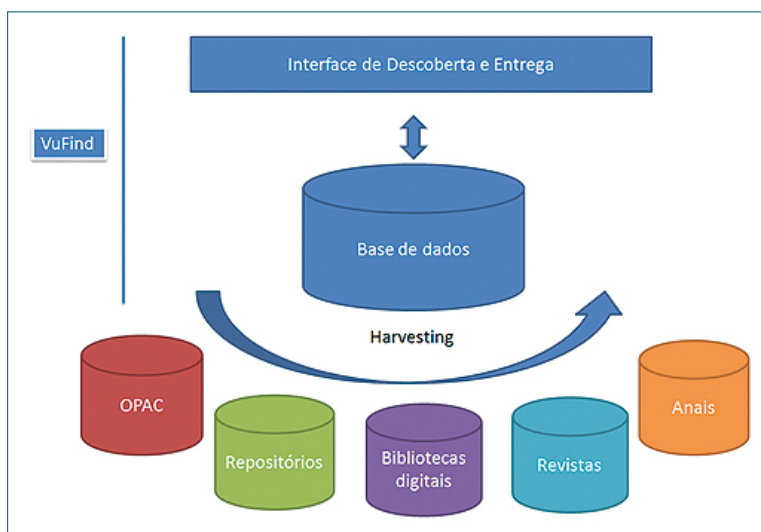


Figura 7 – Estrutura do VuFind  
Fonte: IBICT WIKI (2017).

## CONSIDERAÇÕES FINAIS

O presente capítulo procurou abranger tecnologias para gestão da informação em diversos contextos, visto que não existe uma única ferramenta que atenda a todas as necessidades informacionais de uma instituição. Assim, contempla tecnologias testadas que atendam a automatização de processos específicos, com software livre de custo de licenciamento.

Revela-se que as tecnologias atuam em processos distintos, mas que se complementam no contexto da gestão da informação em uma instituição, principalmente as de ensino e pesquisa. Com isso, automatizam-se tarefas, facilitando a sua execução e apoiando os seus usuários em diversas áreas, contribuindo como um todo à instituição.

Por fim, espera-se contribuir com a discussão sobre tecnologias e gestão da informação, apresentando ferramentas que atendem a atividades específicas. Nota-se que cada vez mais requerem-se tecnologias que não apenas apoiem a execução das atividades, mas que possam contribuir com a gestão, atendendo à dinamicidade do cenário atual.

## REFERÊNCIAS

- ADEMA, J. **Overview of open access models for ebooks in the humanities and social sciences**. Amsterdam: OAPEN, 2010. Disponível em: <<http://project.oapen.org/images/documents/openaccessmodels.pdf>>. Acesso em: 24 jul. 2107.
- AMANTE, M. J.; MARÇAL, B. Dinâmicas de implementação de software open source numa biblioteca universitária: o caso do ISCTE-IUL. In: CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 11., 2012, Lisboa. **Anais...** Lisboa: Associação Portuguesa de Bibliotecários, Arquivistas e Documentalistas, 2012. Disponível em: <<https://repositorio.iscte-iul.pt/handle/10071/3896>>. Acesso em: 23 fev. 2105.
- ANURADHA, K. T.; SIVAKAMINATHAN, R.; ARUN KUMAR, P. Open-source tools for enhancing full-text searching of opacs: use of Koha, greenstone and fedora. **Program**, v. 45, n. 2, p. 231-239, 2011.
- BAPTISTA, A. et al. Comunicação científica: o papel da Open Archives Initiative no contexto do acesso livre. **Encontros Bibli**, n. esp., jan./jun. 2007. Disponível em: <<http://dx.doi.org/10.5007/1518-2924.2007v12nesp1p1>>. Acesso em: 17 mar. 2013.
- BARRETO, A. A oferta e a demanda da informação: condições técnicas, econômicas e políticas. **Ciência da Informação**, v. 28, n. 2., 1999. Disponível em: <<http://revista.ibict.br/ciinf/article/view/847>>. Acesso em: 23 fev. 2015.
- BERGAMASCHI, R.; DUARTE, P. O SEER e o processo de editoração de revistas científicas brasileiras. In: ENCONTRO VIRTUAL DE DOCUMENTAÇÃO EM SOFTWARE LIVRE E CONGRESSO INTERNACIONAL DE LINGUAGEM E TECNOLOGIA ONLINE, 9., 2012, Belo Horizonte. **Anais eletrônicos...** Disponível em: <[http://www.periodicos.letras.ufmg.br/index.php/anais\\_linguagem\\_tecnologia/article/viewFile/1943/3554](http://www.periodicos.letras.ufmg.br/index.php/anais_linguagem_tecnologia/article/viewFile/1943/3554)>. Acesso em: 25 jan. 2015.
- BJÖRK, B. A model of scientific communication as a global distributed information system. **Information Research**, v. 12, n. 2, 2007. Disponível em: <<http://www.informationr.net/ir/12-2/paper307.html>>. Acesso em: 19 jun. 2013.
- BRITO, R. F.; SHINTAKU, M. **Material didático do Curso de OMP**. Brasília: Universidade de Brasília, 2016.
- BUCKLAND, M. Information as thing. **Journal of the American Society for Information Science**, v. 42, n. 5, p. 351-360, 1991.
- CARLOS, S. A. Dez anos de revista estudos interdisciplinares sobre o envelhecimento. **Estudos interdisciplinares Sobre o Envelhecimento**, v. 14, n. 1, 2009. Disponível em: <<http://www.seer.ufrgs.br/RevEnvelhecer/article/view/11907/0>>. Acesso em: 30 maio 2016.
- COHEN, D. **Introducing Omeka**. 2008. Disponível em: <<http://www.dancohen.org/2008/02/20/introducing-omeka/>>. Acesso em: 01 abr. 2016.
- CORRÊA, A. S.; CORRÊA, P. L. P.; SILVA, F. S. C. da. Transparency portals versus open government data: an assessment of openness in Brazilian municipalities. In: ANNUAL INTERNATIONAL CONFERENCE ON DIGITAL GOVERNMENT RESEARCH, 15., 2014, Aguascalientes. **Proceedings...** New York: ACM, 2014. p. 178-185. Disponível em: <<http://dl.acm.org/citation.cfm?id=2612760>>. Acesso em: 18 jan. 2017.

DENTON, W.; COYSH, S. Usability testing of VuFind at an academic library. **Library Hi Tech**, v. 29, n. 2, p. 301-319, 2011.

EGUNJOBI, R.; AWOYEMI, R. Library automation with Koha. **Library Hi Tech News**, v. 29, n. 3, p. 12-15, 2012.

EMANUEL, J. Usability of the VuFind next-generation online catalog. **Information Technology and Libraries**, v. 30, n. 1, p. 44-52, 2011. Disponível em: <<https://ejournals.bc.edu/ojs/index.php/ital/article/view/3044/2666>>. Acesso em: 19 fev. 2016.

GARVEY, D.; GRIFFITH, B. Communication and information processing within scientific disciplines: empirical findings for psychology. **Information Storage and Retrieval**, v. 8, n. 3, p. 123-136, 1972.

HARNAD, S. et al. The access/impact problem and the green and gold roads to open access. **Serials Review**, v. 30, n. 4, p. 310-314, 2004. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0098791307001530>>. Acesso em: 12 jun. 2014.

IBICT WIKI. **VuFind**. Disponível em: <<http://wiki.ibict.br/index.php/Vufind>>. Acesso em: 23 jul. 2017.

JORENTE, M. J. V. et al. O marco civil da internet e a ciência da informação: uma discussão sobre os softwares livres AtoM e Archivematica. **Liinc em Revista**, v. 12, n. 1, 2016. Disponível em: <<http://basessibi.c3sl.ufpr.br/brapci/v/a/22285>>. Acesso em: 05 maio 2016.

KOHA WIKI. Disponível em: <[https://wiki.koha-community.org/wiki/Main\\_Page](https://wiki.koha-community.org/wiki/Main_Page)>. Acesso em: 6 abr. 2015.

KUCSMA, J.; REISS, K.; SIDMAN, A. Using Omeka to build digital collections: the METRO case study. **D-Lib magazine**, v. 16, n. 3/4, 2010. Disponível em: <<http://www.dlib.org/dlib/march10/kucsma/03kucsma.html>>. Acesso em: 06 abr. 2015.

LAMPERT, S. R. Os repositórios DSpace e Archivematica para documentos arquivísticos digitais. **Acervo**, v. 29, n. 2, p. 143-154, nov. 2016. Disponível em: <<http://revista.arquivonacional.gov.br/index.php/revistaacervo/article/view/718>>. Acesso em: 24 fev. 2015.

LIEVROUW, L. A. Communication and the social representation of scientific knowledge. **Critical Studies in Mass Communication**, v. 7, n. 1, p. 1-10, mar. 1990.

LYNCH, C. A. Institutional repositories: essential infrastructure for scholarship in the digital age. **Portal: Libraries and the Academy**, v. 3, n. 2, p. 327-336, 2003. Disponível em: <<https://muse.jhu.edu/article/42865/pdf>>. Acesso em: 06 abr. 2015.

MACÊDO, D.; SHINTAKU, M.; BRITO, R. Dublin core usage for describing documents in Brazilian government digital libraries. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2015, São Paulo. **Anais eletrônicos**... p. 129-135. Disponível em: <<http://dcpapers.dublincore.org/pubs/article/view/3768>>. Acesso em: 25 abr. 2016.

MANUAL de gestão de unidades de informação. Curitiba: TECPAR; Brasília: IBICT, 1997

MAQUILÓN SANCHES, J.; LILLO HIDALGO, V.; MIRETE RUIZ, A. B. La edición electrónica y la gestión de trabajos académicos con Open Conference Systems (OCS). **REIFOP**, v. 14, n. 1, p. 325-344, 2011. Disponível em: <[http://www.aufop.com/aufop/uploaded\\_files/articulos/1301819863.pdf](http://www.aufop.com/aufop/uploaded_files/articulos/1301819863.pdf)>. Acesso em: 06 abr. 2015.

MARCHIORI, P. Z. A ciência e a gestão da informação: compatibilidades no espaço profissional. **Ciência da Informação**, v. 31, n. 2, out. 2002. Disponível em: <<http://revista.lbict.br/ciinf/article/view/962>>. Acesso em: 24 fev. 2016.

MARTINS, A. et al. O Museu da Universidade de Aveiro: uma viagem para além do espaço físico. In: ACTAS DO CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 12., 2015, Évora, Portugal. **Anais eletrônicos...** p. 1-12. Disponível em: <[http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/1383/pdf\\_83](http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/1383/pdf_83)>. Acesso em: 25 abr. 2016.

MIRANDA, A. C. D., MORAES, M. H. M.; BARROS, J. L. Portal de Periódicos da FURG, acesso aberto e ampliação da produção científica institucional. In: EUSEER, 1., 2010, Florianópolis. **Anais eletrônicos...** Disponível em: <<http://euseer.lbict.br/index.php/iieuseer/euseer1/paper/view/34>>. Acesso em: 25 abr. 2016.

MURAKAMI, T.; FAUSTO, S. Panorama atual dos repositórios institucionais das instituições de ensino superior no Brasil. **InCID: Revista de Ciência da Informação e Documentação**, v. 4, n. 2, p. 185-201, 2013. Disponível em: <<http://www.revistas.usp.br/incid/article/view/69327>>. Acesso em: 24 out. 2016.

OLETO, R. R. Percepção da qualidade da informação. **Ciência da Informação**, v. 35, n. 1. 2006. Disponível em: <<http://revista.lbict.br/ciinf/article/view/1153/1316>>. Acesso em: 1 dez. 2015.

PÉREZ, T. Exposiciones digitales y reutilización: aplicación del software libre Omeka para la publicación estructurada. **MÉI: Métodos de Información**, v. 2, n. 2, p. 29-46, 2011. Disponível em: <<https://dialnet.unirioja.es/servlet/articulo?codigo=3720111>>. Acesso em: 24 out. 2016.

PIRES, E. A. N. A.; KLEBERSSON, M. Utilização do Latindex como ferramenta de avaliação de periódicos eletrônicos científicos: uma análise da Biblionline. In: ENCONTRO NACIONAL DE ESTUDANTES DE BIBLIOTECONOMIA, DOCUMENTAÇÃO, CIÊNCIA DA INFORMAÇÃO E GESTÃO DA INFORMAÇÃO ESCOLA DE CIÊNCIA DA INFORMAÇÃO, 35., 2012, Belo Horizonte. **Anais eletrônico...** Belo Horizonte: Universidade Federal de Minas Gerais, 2012. Disponível em: <<http://rabci.org/rabci/node/412>>. Acesso em: 17 ago. 2016.

SCHEINFELDT, T. Omeka: open source web publishing for research, collections and exhibitions. **Open Source Business Resource**, dec., 2008. Disponível em: <<https://timreview.ca/article/211>>. Acesso em: 01 dez. 2016.

SCHIESSL, I. T. et al. Koha: um sistema integrado de gerenciamento de bibliotecas. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 19., 2016, Manaus. **Anais eletrônicos...** Disponível em: <<http://periodicos.ufam.edu.br/anaissnbu/article/view/3329>>. Acesso em: 24 fev. 2017.

SHINTAKU, M.; BRITO, R.; FLEURY, A. **SOAC/OCS para gerentes gerais**. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2014. Disponível em: <<http://labcoat.lbict.br/portal/wp-content/uploads/2015/03/Item-12-Digital-2.pdf>>. Acesso em: 24 out. 2015.

SHINTAKU, M.; BRITO, R.; CARVALHO NETO, S. A avaliação dos portais de revistas brasileiros implementados com o SEER/OJS por meio do levantamento da indexação pelo Latindex e SciELO. **Informação & Sociedade: estudos**, v. 24, n. 2, 2014. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/18671>>. Acesso em: 24 out. 2015.



- SHINTAKU, M.; MACEDO, D. J. Sobre o VuFind, software livre de descoberta e entrega. **De olho na CI.**, v. 6, n. 1, 2015. Disponível em: <<http://www.deolhonaci.com/news/sobre-o-vufind-software-livre-de-descoberta-e-entrega/>>. Acesso em: 24 out. 2016.
- SHINTAKU, M.; VIDOTTI, S. Bibliotecas e repositórios no processo de publicação digital. **BIBLOS**, v. 30, n. 1, p. 61-80, 2016. Disponível em: <<https://www.seer.furg.br/biblos/article/view/5762/3993>>. Acesso em: 24 fev. 2017.
- SON, Yun-hee et al. Platform design for data sharing. In: PARK, James J. et al. (Ed.). **Frontier and innovation in future computing and communications**. Springer: Nova York, 2014. p. 447-442. Disponível em: <[http://link.springer.com/chapter/10.1007/978-94-017-8798-7\\_54](http://link.springer.com/chapter/10.1007/978-94-017-8798-7_54)>. Acesso em: 19 jan. 2017
- SUTRADHAR, B. **One day national workshop on web-scale discovery services**: single window access to library e-resources. 2014. Disponível em: <<http://www.library.iitkgp.ernet.in/sites/bookFair/wkp/OneDayNationalWorkshop.pdf>>. Acesso em: 24 fev. 2017.
- TANSLEY, R. et al. The DSpace institutional digital repository system: current functionality. In: JOINT CONFERENCE ON DIGITAL LIBRARIES, 3., 2003, USA. **Proceedings...** p. 87-97. Disponível em: <[https://dspace.mit.edu/bitstream/handle/1721.1/26705/Tansley\\_2003\\_The.pdf?sequence=1](https://dspace.mit.edu/bitstream/handle/1721.1/26705/Tansley_2003_The.pdf?sequence=1)>. Acesso em: 24 out. 2015.
- THIAGARAJAN, B. **Open monograph press installation manual**. India: Select Words, 2013.
- VAN GARDEREN, P. Archivematica: using micro-services and open-source software to deliver a comprehensive digital curation solution. In: INTERNATIONAL CONFERENCE ON PRESERVATION OF DIGITAL OBJECTS, 7., 2010, Vienna, Austria. **Proceedings...** Disponível em: <<http://www.ifs.tuwien.ac.at/dp/ipres2010/papers/vanGarderen28.pdf>>. Acesso em: 24 out. 2015.
- VAUGHAN, J. Web scale discovery services. **Library Technology Reports**, v. 47 n.1, jan. 2011.
- WILLINSKY, J. Toward the design of an Open Monograph Press. **Journal of Electronic Publishing**, v. 12, n. 1, 2009. Disponível em: <<http://quod.lib.umich.edu/jep/3336451.0012.103?view=text;rgn=main>>. Acesso em: 15 nov. 2015.
- YANG, S. Q.; HOFMANN, M. A. The next generation library catalog: a comparative study of the OPACs of Koha, Evergreen, and Voyager. **Information Technology and Libraries**, v. 29, n. 3, p. 141, 2010. Disponível em: <<https://ejournals.bc.edu/ojs/index.php/ital/article/viewFile/3139/27>>. Acesso em: 15 nov. 2015.



# POLÍTICAS EM REPOSITÓRIOS DIGITAIS: DAS DIRETRIZES À IMPLEMENTAÇÃO

Emanuelle Torino





## INTRODUÇÃO

A nova dinâmica na produção da informação, impulsionada pelo avanço das tecnologias de informação e comunicação (TIC), ocasionou a conhecida explosão informacional, que altera o contexto da recuperação e do uso da informação.

É crescente ainda a possibilidade de formatos de arquivo, as formas de comunicação e conexão entre indivíduos, o desenvolvimento de mídias para a criação e o acesso a conteúdos, as possibilidades de compartilhamento e de armazenamento de arquivos e tantas outras questões que permeiam o universo da informação.

Nesse contexto, o ciclo vital da informação precisa ser planejado, de forma a auxiliar o autor de um documento, no que tange às maneiras de tornar o conteúdo visível, acessível e preservado, além de e permitir que autor e leitor tenham assegurado o acesso e o uso do material.

Tratar este emaranhado de conexões e de conteúdos e apresentá-los de forma organizada, considerando a relevância do documento, é o desafio dos profissionais da informação.

Em meados da primeira década de 2000, ganharam força as primeiras discussões acerca do Movimento de Acesso Aberto (OA) que discutiu, na Budapest Open Access Initiative (BOAI) (2002), a convergência entre a prática comum entre os pesquisadores (de publicar gratuitamente os resultados de suas pesquisas visando ao avanço da ciência) e a possibilidade de expandir o acesso a estas publicações arbitradas por pares, de forma gratuita e irrestrita a qualquer interessado utilizando a internet.

A BOAI definiu duas estratégias necessárias para atingir a meta do acesso aberto:

- a) a alteração na gestão dos periódicos, livres de custos de assinaturas ou taxas de acesso e com o uso do *copyright* como garantia dos direitos do autor e da fonte publicadora, e não mais como restrição ao acesso. A chamada **via dourada** busca privilegiar a publicação em revistas de acesso aberto, cujos direitos de uso são visíveis ao leitor e com a adoção de um protocolo de interoperabilidade de metadados (*Open Archive Initiative – Protocol for Metadata Harvesting* – OAI-PMH) capaz de coletar e distribuir os metadados a outras fontes na internet;

- b) a disponibilização de repositórios eletrônicos abertos, mantidos por universidades ou instituições de pesquisa, nos quais os autores a elas vinculados devem depositar a sua produção, ao que se conhece como **via verde**. Os repositórios também é utilizam-se do protocolo OAI-PMH.

Nesse sentido, para quebrar as barreiras de acesso, é necessário o comprometimento coletivo, de instituições e de indivíduos, em prol do acesso aberto e da ampla disseminação dos resultados de pesquisa.

Tais discussões e iniciativas se fizeram conhecidas por bibliotecários e aos poucos ganharam força junto a editores de periódicos, gestores de agências de fomento e de instituições acadêmicas e de pesquisa.

Em 2012, passados 10 anos da primeira BOAI, considerando que os objetivos não foram alcançados totalmente e as experiências adquiridas, foram adicionadas novas recomendações para os próximos 10 anos, no que tange a: políticas, licenciamento e reutilização, infraestrutura e sustentabilidade, promoção e coordenação (BUDAPEST OPEN ACCESS INITIATIVE, 2012).

Tal fato faz com que um cenário complexo passe a ser visto sob uma nova perspectiva. Sabe-se que longo ainda é o caminho a ser trilhado para que os grandes editores científicos alterem o modelo de negócio das publicações com vistas ao acesso aberto. Contudo, desde o início das discussões, ocorreram avanços significativos, que podem ganhar força com o passar de mais alguns anos.

A disponibilização de repositórios digitais para a gestão da informação científica ou acadêmica é uma realidade em cenário internacional. Tal realidade ganha força também em muitas instituições no Brasil e, para que o desenvolvimento deste serviço ocorra de forma satisfatória, é necessária a atuação embasada em uma política.

O presente capítulo pretende discutir o gerenciamento, o acesso e o uso da produção disponível em repositórios digitais, considerando aspectos relacionados às suas políticas.

## **GESTÃO DE REPOSITÓRIOS DIGITAIS**

Os repositórios digitais (RDs) são sistemas de informação abertos e interoperáveis destinados à gestão da informação científica e acadêmica, capazes de armazenar arquivos de diversos formatos, constituindo-se em vias alternativas de comunicação científica e ampliação de visibilidade da produção.

Para Camargo e Vidotti (2011), os RDs foram instituídos com a finalidade de preservação de memória e visibilidade institucional. Inicialmente atinham-se à produção científica, mas podem também ser desenvolvidos com finalidade ad-



ministrativa. Afirmam ainda que, por se tratar de temática recente, é passível de sofrer alterações conceituais.

O tipo de repositório digital é determinado pela aplicação e pelos objetivos aos quais se destina, destacando-se os repositórios institucionais e os temáticos.

Para Leite (2009, p. 21), um repositório institucional (RI) constitui:

um serviço de informação científica - em ambiente digital e interoperável - dedicado ao gerenciamento da produção intelectual de uma instituição. Contempla, por conseguinte, a reunião, armazenamento, organização, preservação, recuperação e, sobretudo, a ampla disseminação da informação científica produzida na instituição.

O repositório temático (RT) trata a produção intelectual por área do conhecimento e está voltado a comunidades científicas específicas (LEITE, 2009).

Ao definir RI, Leite (2009) elenca o tratamento dado à informação por parte dos repositórios digitais, tratamento esse intimamente ligado ao processo de gestão da informação desenhado por Davenport (1998), constituído por determinação das exigências, obtenção, processamento, distribuição e utilização. A compilação deste processo com as teorias de outros autores da área (como McGEE; PRUSAK, 1994; STAIR, 1998; CASSARRO, 1999; DAVENPORT, 1998) levou à síntese elaborada por Moraes e Escrivão Filho (2006), explicitada na Figura 1.

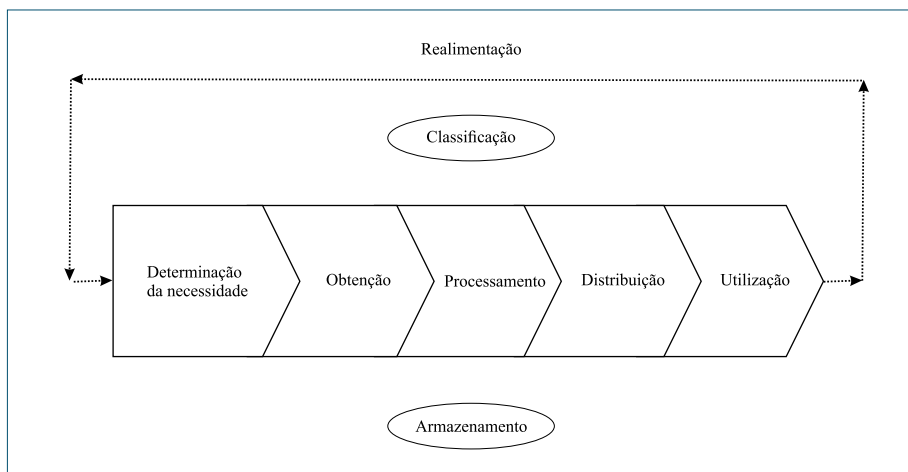


Figura 1 – O processo de gerenciamento da informação

Fonte: Moraes e Escrivão Filho (2006).

O processo de gerenciamento da informação permeia a atividade dos repositórios digitais, desde o seu planejamento, implementação e desenvolvimento, e está intimamente ligado à sua gestão.

Na determinação de necessidades, se concentra o processo base do RD como serviço de informação, e nele está contida a definição de diretrizes que nortearão a atividade, que devem estar descritas em uma política.

É fundamental para a gestão do RD que ele esteja alinhado à instância à qual está vinculado, e que reflita também seus interesses. Se tratarmos de um RI, a vinculação fica mais evidente, contudo, mesmo o RT possui uma linha norteadora definida pela sociedade científica à qual se destina.

Isso posto, identifica-se o papel a ser desenvolvido pelo gestor do RD, tornando-lo peça central na sua comunidade, a partir do alinhamento de interesses que determinarão os objetivos do serviço. Atrelado a isso, deve haver especial atenção ao cumprimento de diretrizes necessárias para que o sistema atenda aos preceitos do OA e às diretrizes nacionais e internacionais para cumprir adequadamente o seu papel de fornecer acesso aberto, livre e irrestrito à informação, bem como ampliar sua visibilidade.

## **POLÍTICAS EM REPOSITÓRIOS DIGITAIS**

A elaboração de políticas pode ser destacada como uma das etapas mais relevantes do planejamento de um repositório, uma vez que nortearão não apenas a implantação, mas a forma e a velocidade do seu desenvolvimento.

Embora as políticas de informação sejam vistas como tema emergente após a Segunda Guerra Mundial, como afirmam Jardim, Silva e Nharreluga (2009), no Brasil, a ausência de uma política pública acerca da temática leva cada instituição e cada repositório a estabelecer definições próprias.

Em geral, na etapa de planejamento, muitos são os questionamentos e, mesmo depois do repositório em funcionamento, ainda restam incertezas. No que tange às políticas: quais políticas são necessárias? De acesso aberto, de informação, de conteúdo, de funcionamento, de submissão?

Rodrigues (2009) disponibiliza um conjunto de práticas, modelos e propostas de implementação de política de acesso aberto visando apoiar instituições e agências de fomento. Para o autor, a implementação de uma política de acesso aberto requer vontade política, definição de responsabilidades de execução e controle e, sobretudo, a existência de um repositório.

A literatura da área de ciência da informação traz algumas conceituações relevantes acerca de política de informação, a exemplo da apresentada por Silva (1991, p. 12): “numa visão ideal, uma política de informação deve ser um instrumento que integre a sociedade aos avanços científicos e tecnológicos, de forma participativa”.



Tal afirmação reforça a necessidade de uma política nacional de informação, que estabeleça diretrizes para a atuação das instituições quanto aos seus fluxos informacionais, formas de abertura e acesso.

Para Hernon e Relyea (1991 apud JARDIM; SILVA; NHARRELUGA, 2009, p. 9), a política de informação é:

um conjunto de princípios, leis, diretrizes, regras, regulamentos e procedimentos interrelacionados que orientam a supervisão e gestão do ciclo vital da informação: a produção, coleção, organização, distribuição/disseminação, recuperação e eliminação da informação. Política de informação compreende o acesso à, e uso da informação.

A política de funcionamento, por sua vez, “deve refletir as decisões tomadas ao longo do planejamento do repositório. É recomendável que esta política esteja em concordância com aquelas já vigentes na biblioteca e na instituição” (LEITE et al., 2012, p. 10). Segundo esses autores, deve compreender: definição do serviço; formação da equipe; prazo definido para o depósito no repositório; tipo de material aceito; responsabilidade pelo depósito e fluxo de trabalho.

Ao planejar uma política, seja para implementação, seja para atualização, é imprescindível ter em mente que “a essência da política de informação é formada por que faz escolhas e pelas consequências que essa escolha determina” (DAVENPORT, 1998, p. 91). Ao que Tomaél e Silva (2007) complementam: “uma política de informação cobre os objetivos, diretrizes, práticas e intenções organizacionais que servem para fortalecer as decisões locais”.

Dessa forma, para que a disponibilização e o funcionamento do RD estejam formalmente documentados, podem ser definidos um ou mais documentos que apresentem as suas diretrizes e linhas norteadoras. Neste capítulo, os elementos necessários serão descritos e atrelados sempre à política.

Nesse sentido, a Budapest Open Access Initiative (2012) recomenda que:

Todas as instituições de ensino superior devem ter uma política que assegure que versões revisadas por pares, de todos os futuros artigos científicos da autoria dos seus membros, sejam depositadas no repositório designado pela instituição.

É imprescindível que a política estabeleça: objetivo; equipe, instâncias e atribuições; mandato e mecanismos de acompanhamento; arquitetura da informação; gestão de coleções; metadados; tipologias, formatos e tamanho de arquivos; direito autoral e embargo (formas de acesso); formas de povoamento; fluxo de trabalho; tratamento da informação; preservação digital e atualização.

## OBJETIVO

Ao planejar e disponibilizar qualquer serviço de informação, é necessário que haja um alinhamento aos interesses da(s) instância(s) à qual o serviço está atrelado.

Na definição de uma política para repositório, é imprescindível que se estabeleça o objetivo da própria política e que haja interlocução com os demais documentos institucionais e da biblioteca, que, em geral, é o órgão responsável pela gestão e tratamento da informação científica e acadêmica e, portanto, pelo RD.

Sempre que possível e desejado pela instituição mantenedora, é recomendável que a política mantenha como objetivo o acesso aberto e que incentive e encoraje os pesquisadores à publicação em periódicos abertos, utilizando-se de uma licença de direito autoral.

Além disso, deve apresentar o objetivo do RD, sua área de atuação e posicionamento como ferramenta de gestão e ampliação da visibilidade da informação científica e acadêmica institucional.

## EQUIPE, INSTÂNCIAS E ATRIBUIÇÕES

A disponibilização de um RD requer interesse e apoio da instituição mantenedora, o que torna necessário o envolvimento de instâncias superiores que atuem na tomada de decisão institucional, para reforçar e envolver a comunidade nos objetivos e benefícios do repositório, atribuindo a este o papel de ferramenta institucional de gestão da informação científica e acadêmica.

Considerando a estrutura organizacional, podem ser envolvidas diferentes instâncias, às quais devem ser imputadas atribuições específicas, definidas pela área de atuação.

Por exemplo, em uma universidade, podem ser estabelecidas as responsabilidades da reitoria, das pró-reitorias, dos departamentos acadêmicos, do setor de recursos humanos, da área de informática, da biblioteca, dos pesquisadores, dos servidores e dos alunos, enquanto em uma instituição de pesquisa é importante que se estabeleçam as atribuições da direção, da área de informática, da biblioteca, dos pesquisadores. Esta cadeia de envolvidos é que fortalece a presença e a atuação do repositório, atendendo sempre aos preceitos estabelecidos pela política.

À equipe compete exercer as atividades vitais de manutenção de software, parametrização, padronização e funcionamento, seguindo aos preceitos estabelecidos pela política. Por isso é imprescindível estabelecer as equipes de trabalho e suas atribuições.

## Comitê Gestor

Para a atuação diretiva é necessário que haja um comitê gestor formalizado, do qual façam parte representantes de todas as áreas que possuam interlocução com a produção de conteúdos gerenciáveis pelo RD.

Este comitê é responsável pela atuação integrada aos interesses institucionais e por ações como:

- a) estabelecer diretrizes gerais de funcionamento do repositório, atendendo às necessidades e objetivos institucionais, bem como sua relação com a sociedade;
- b) manter atualizada a política que abranja todos as diretrizes necessárias ao funcionamento adequado do repositório;
- c) disseminar o serviço e a política;
- d) acompanhar o desenvolvimento do RD, bem como o cumprimento da política por todos os atores;
- e) deliberar sobre questões relativas ao serviço;
- f) propor melhorias no processo.

## Gestor do Repositório

A gestão de um RD deve ser desempenhada por profissional capacitado, que compreenda os fluxos e estoques informacionais, bem como suas formas de tratamento, visando à recuperação da informação e à visibilidade web. Além disso, é necessário que possua habilidades humanísticas, de forma a envolver os diferentes atores e departamentos na consecução dos objetivos a ele estabelecidos pela política.

Suas atividades consistem em:

- a) gerir recursos humanos, financeiros e tecnológicos envolvidos com o RD;
- b) planejar o desenvolvimento do serviço, visando melhorias constantes;
- c) elaborar e monitorar fluxos de informação e o povoamento;
- d) gerenciar o funcionamento do serviço, bem como as atividades da equipe executiva;
- e) elaborar relatórios necessários à instituição mantenedora;
- f) mediar a relação entre os diferentes atores;
- g) realizar testes necessários às novas implementações;
- h) atuar como administrador do software;
- i) criar perfis ou grupos de usuários e parametrizar o uso do sistema;

- j) criar a estrutura de organização da informação do repositório. Caso haja a definição de administradores por coleção, esta função pode ser delegada a eles;
- k) estabelecer, em conjunto com a equipe executiva, o perfil de aplicação de metadados;
- l) definir, em conjunto com a equipe executiva e de informática, ações necessárias à preservação digital;
- m) proporcionar capacitação à equipe;
- n) deliberar sobre quaisquer questões a ele encaminhadas, relativamente ao serviço e, quando necessário, recorrer ao Comitê Gestor;
- o) demais ações pertinentes.

### Equipe Executiva

O funcionamento adequado do serviço, enquanto unidade executiva, exige a formação de uma equipe dedicada exclusivamente à atividade, preferencialmente multidisciplinar e que conte com profissionais de informação e de informática, que podem ser apoiados por outros profissionais por eles supervisionados.

Aos profissionais de informação compete:

- a) auxiliar no cumprimento da política;
- b) adotar os procedimentos administrativos e operacionais necessários ao funcionamento do repositório;
- c) utilizar padrões nacionais e internacionais de metadados;
- d) empregar a forma adotada para o preenchimento dos campos, inclusive no que tange à padronização de autoridades;
- e) descrever os itens, quando estabelecido no fluxo de trabalho;
- f) revisar as submissões, quando estabelecido no fluxo de trabalho;
- g) certificar-se da permissão para disponibilização dos itens (direito autoral e uso de licenças);
- h) proporcionar capacitação aos usuários;
- i) supervisionar as atividades dos auxiliares;
- j) demais atividades pertinentes.

Aos profissionais de informática compete:

- a) auxiliar no cumprimento da política;

- b) manter o software disponível, acessível, seguro e atualizado, considerando a última versão disponível e estável;
- c) promover estudos e recomendar procedimentos e padrões técnicos para a segurança do serviço;
- d) propor implementações ao código fonte, quando se fizer necessário, de forma a contribuir com a comunidade de desenvolvedores;
- e) auxiliar, quando definido, nas ações relacionadas à coleta e depósito automático de registro;
- f) customizar e administrar os softwares necessários ao serviço;
- g) manter uma política de backup;
- h) supervisionar as atividades dos auxiliares;
- i) demais atividades pertinentes.

Aos profissionais de marketing compete:

- a) auxiliar no cumprimento da política;
- b) desenvolver plano de divulgação do serviço, bem como materiais de divulgação;
- c) propor identidade visual para fixar a marca;
- d) orientar a equipe no direcionamento de mecanismos para cobertura de necessidades específicas dos usuários;
- e) supervisionar as atividades dos auxiliares;
- f) demais atividades pertinentes.

A todas as atividades descritas podem ser somadas, ou delas excluídas, atribuições considerando as características institucionais e do RD, registrando-as na política.

A forma de povoamento e o fluxo de trabalho estabelecido é que determinarão a atuação dos atores nas atividades do repositório. O papel do autor, por exemplo, poderá ser fortemente alterado pelo mecanismo adotado para o povoamento. Independentemente disso, tanto a instituição quanto o comitê gestor e demais membros da equipe precisam trabalhar para que os autores dos conteúdos compreendam a relevância e a necessidade de disponibilização no repositório.

Há ainda de se pensar no papel das agências de fomento e, da própria instituição quando do fomento a novos projetos e pesquisas, atrelando a disponibilização dos resultados no RD como forma de garantir acesso perpétuo aos resultados de forma rápida, organizada e segura.

## MANDATO E MECANISMOS DE ACOMPANHAMENTO

Para assegurar o fluxo contínuo de depósitos, a política deve estabelecer um mandato de depósito obrigatório da produção dos pesquisadores vinculados à instituição mantenedora. A Budapest Open Access Initiative (2012) recomenda que:

Todas as políticas [...] devem requerer o depósito num repositório adequado entre a data de aceite para publicação e a data de publicação. Os metadados devem ser depositados logo que estejam disponíveis e devem ser AA a partir do momento do depósito. O texto integral deve ser publicado em AA logo que o repositório tenha permissão para o fazer.

Leite (2009) afirma que: “os repositórios com mais sucesso no que diz respeito ao seu povoamento têm sido aqueles cujas instituições estabeleceram políticas de depósito obrigatório [...] fator primordial para que as taxas de depósito sejam consideravelmente aumentadas”.

Vale destacar que o mandato deve cobrir toda a produção gerenciada pelo repositório e as tipologias documentárias por ele aceitas. Deve estabelecer ainda que a disponibilização ocorra o mais breve possível, atendendo às condições imputadas pela fonte publicadora. É relevante mencionar que, caso haja algum impedimento para a disponibilização imediata do material, é possível inseri-lo com restrição de acesso, com a possibilidade de acesso imediato ou restrito aos metadados.

É importante que as agências de fomento também estabeleçam mandatos para as produções oriundas de projetos por elas apoiados.

Todas as agências financiadoras de pesquisa, públicas ou privadas, devem ter uma política assegurando que versões com revisão por pares de todos os futuros artigos resultados de pesquisa financiada sejam depositados num repositório adequado e disponibilizados em AA logo que possível (BUDAPEST OPEN ACCESS INITIATIVE, 2012).

Além do mandato, devem ser definidas formas de acompanhamento que assegurem que toda a produção seja disponibilizada. Uma boa prática para universidades é atrelar a disponibilização da produção dos servidores às progressões funcionais, prática esta também recomendada pela Budapest Open Access Initiative (2012): “as universidades que possuam um repositório institucional devem exigir o depósito no repositório de todos os artigos científicos que devam ser considerados na promoção daquele acadêmico, no seu curso profissional ou em quaisquer outras formas avaliação interna”.

## ARQUITETURA DA INFORMAÇÃO

Os repositórios digitais devem estar centrados no usuário, considerando a premissa da visibilidade dos conteúdos e do acesso à informação. Para tanto, é necessário estabelecer um fluxo efetivo da informação disponível em seu ambiente, organizando-a para favorecer o acesso.

A essa organização, Wurman (1996) chamou de arquitetura de informação, tida como um mapa que permite ao usuário encontrar caminhos para a construção de conhecimentos.

Para Davenport (1998, p. 222), “boas arquiteturas são essenciais para processos informacionais eficientes, em especial quando lidam com acesso, obtenção e distribuição de dados”.

Corroboram Camargo e Vidotti (2011, p. 24), que definem arquitetura da informação como uma área do conhecimento que atua com “aspectos informacionais, estruturais, navegacionais, funcionais e visuais de ambientes informacionais digitais, [...] a fim de auxiliar no desenvolvimento e no aumento da usabilidade de tais ambientes e seus conteúdos”.

Nesse sentido, no repositório, é necessário planejar criteriosamente o caminho a ser seguido pelo usuário para acessar a informação de que necessita.

Os sistemas estruturantes propostos por Morville e Rosenfeld (2006): organização, rotulagem, navegação, busca, metadados, vocabulário controlado e tesouros devem ser considerados neste planejamento.

Em geral, a arquitetura da informação apresentada está baseada no cenário organizacional refletindo sua estrutura ou elementos que a caracterizam. Contudo, é necessário se atentar para que os rótulos apresentados estejam imbuídos de valor para usuários externos à instituição e sejam diferentes entre si, de modo a serem interpretados por humanos e máquinas.

Vale ressaltar que os campos de metadados disponíveis podem ser utilizados como filtro de apresentação de resultados, tornando-se formas alternativas de acesso ou facetas. Há de se destacar, contudo, que a interface do repositório deve ser limpa e atrativa para facilitar a navegação e o acesso à informação.

## GESTÃO DE COLEÇÕES

O software mais utilizado para a disponibilização de repositórios é o DSpace. Sua estrutura hierárquica permite a organização da informação em comunidades, subcomunidades e coleções, os itens e arquivos ficam disponíveis no último nível.

A política de informação deve prever como será apresentada esta estrutura. O mais usual é representar a estrutura organizacional da instituição mantenedora, mas nada impede que outro arranjo seja adotado, a exemplo da organização por tipo de documento.

Nesse sentido, é relevante salientar a necessidade de evitar a duplicidade de nomenclatura na estrutura, a fim de favorecer a coleta pelos mecanismos de busca.

Definida a estrutura, é possível determinar o repositório terá apenas um administrador ou se cada comunidade será gerenciada por um indivíduo ou grupo. Caso a opção seja a adoção de administradores específicos para as comunidades, é imprescindível que haja diretrizes norteadoras que tragam unidade à forma de atuação.

Ao administrador de comunidade compete:

- a) auxiliar no cumprimento da política;
- b) criar estruturas à comunidade;
- c) definir usuários ou grupos de usuários atrelados às coleções presentes na comunidade;
- d) estabelecer, quando necessário, políticas para a coleção, em consonância com a política do repositório;
- e) estabelecer o fluxo de trabalho, de acordo com a política do RD e com políticas próprias da coleção, caso haja;
- f) demais atividades pertinentes.

Ao gestor do repositório ou da comunidade competem, além de executar funções no software, o desenvolvimento e o povoamento da estrutura de forma organizada e atendendo aos preceitos estabelecidos pela política do RD.

## METADADOS

O processo de descrição da informação se faz utilizando-se de metadados. Para Corrado e Jaffe (2014, tradução nossa), “metadados são elementos utilizados para descrever recursos com o propósito de recuperação, gestão de direitos e preservação”. Os metadados auxiliam os usuários na identificação, autenticação e contextualização dos dados, conjunto de dados e outros recursos digitais, estabelecendo relações entre eles, bem como definem permissões e requisitos técnicos para acesso e uso dos objetos, continuam os autores.

Tradicionalmente, o esquema mais utilizado é o *Machine Readable Catalogin* (MARC), porém, com o advento dos sistemas de informação na web, outros se tornaram conhecidos, a exemplo do *Dublin Core* (DC), *Learning Object Metadata*



(LOM), *Metadata Object Description Standard* (MODS), *Metadata Encoding & Transmission Standard* (METS), entre outros.

Vale destacar que cada esquema de metadados possui uma aplicação específica, considerando o objeto a ser descrito e o contexto.

A utilização de um esquema normalizado é fundamental para a interoperabilidade entre sistemas e, conseqüentemente, para a visibilidade dos conteúdos descritos. Nesse sentido, a Budapest Open Access Initiative (2012) assevera que “devemos ajudar na recolha, organização e disseminação de metadados de AA em formatos normalizados para todas as publicações, novas ou antigas, incluindo as publicações não AA”.

Considerando os objetivos do repositório e as especificidades dos materiais a serem descritos, é necessário elencar o conjunto de metadados necessários para a descrição dos conteúdos e desenvolver um perfil de aplicação de metadados, que consiste na reunião de elementos selecionados de um ou mais esquemas de metadados combinados, atrelando-os às tipologias aceitas, além, de determinar a forma de uso.

É importante determinar na política quais os esquemas de metadados utilizados, estabelecer os requisitos de descrição para cada elemento, as formas de padronização e de normalização, visando à qualidade sintática e semântica.

## TIPOLOGIAS, FORMATOS E TAMANHO DE ARQUIVOS

Os repositórios surgiram como ferramentas alternativas de comunicação científica, de forma que, inicialmente, os conteúdos depositados eram estritamente científicos: artigos, trabalhos publicados em eventos, livros e capítulos de livros. A estes conteúdos uniram-se as dissertações e teses, por sua criteriosa avaliação.

Mais tarde, passaram a ser incorporadas outras tipologias, como: documentos acadêmicos, recursos educacionais abertos, áudios, vídeos, apresentações, apostilas, relatórios e, em alguns casos, documentos administrativos.

Nesse sentido, a Budapest Open Access Initiative (2012) afirma:

Encorajamos a experimentação com novas formas de “artigo” e “livro” científicos, nos quais os textos estão integrados de formas úteis com os dados em que se baseiam, elementos multimídia, código executável, literatura relacionada e comentários dos utilizadores.

O que de fato determina os tipos aceitos são os objetivos do repositório. Desta forma, a política deve elencar um conjunto de tipologias aceitas.

Além disso, considerando a preservação digital e o acesso ao conteúdo ao longo do tempo, é necessário estabelecer os formatos de arquivos passíveis de disponibilização, recomendando sempre os que asseguram tais características.

Quanto ao tamanho dos arquivos, cabe ao gestor, em conjunto com a equipe de informática, planejar a estrutura necessária para o repositório no que tange à capacidade do servidor, de *storage* e de memória. Tendo em vista este planejamento, é possível determinar se haverá limitação ao tamanho dos arquivos, sobretudo os não-textuais; e, determinar as necessidades de expansão de infraestrutura.

## DIREITO AUTORAL E FORMAS DE ACESSO

Importante característica de um repositório digital é o fato de disponibilizar informações publicadas em outras fontes, como editoras, periódicos, eventos e outros.

Dessa forma, durante o tratamento da informação, é necessário se atentar para as implicações legais impostas à obra analisada.

Do ponto de vista da política, é relevante que se estabeleça institucionalmente a forma de trabalhar com as publicações a serem depositadas, considerando a titularidade do direito autoral e a possibilidade de uso de licenças abertas, atentando-se sempre à fonte original de publicação e à legislação vigente.

A esse respeito, a Budapest Open Access Initiative (2012) estabelece que:

Os repositórios AA dependem geralmente de permissões de terceiros, como autores ou editores, e raramente estão em posição de requerer licenças abertas. No entanto, os decisores políticos que estejam em posição de direcionar depósitos nos repositórios devem exigir, sempre que possível, licenças abertas, preferencialmente CC-BY.

Considerando as questões legais, pode-se definir a forma de disponibilização: acesso aberto e imediato ao conteúdo completo, acesso ao conteúdo após um prazo pré-estabelecido (embargo), acesso restrito a uma comunidade específica de usuários ou acesso restrito.

Outra questão relevante é a definição institucional da adoção de licenças abertas para as publicações por elas chanceladas, publicadas por editoras, periódicos e eventos institucionais e, até mesmo dissertações e teses.

A instituição pode, ainda, a seu critério, estabelecer diretrizes que incentivem os pesquisadores a ela vinculados a publicar seus conteúdos em editoras e periódicos de acesso aberto.

Vale destacar que a disponibilização dos conteúdos em acesso aberto favorece a visibilidade do conteúdo e, que as instituições, agências de fomento e RDs devem, sempre que possível favorecer essa conduta.

## FORMAS DE POVOAMENTO

As recomendações da Budapest Open Access Initiative (2002) apresentavam o autoarquivamento como forma de povoamento dos repositórios. Nessa modalidade, compete ao autor identificar a condição de disponibilização permitida pela fonte publicadora e efetuar o depósito do conteúdo no repositório.

Na fase de implantação de repositórios no Brasil, uma estratégia adotada por alguns gestores foi o chamado depósito mediado, realizado por profissionais vinculados às bibliotecas e aos repositórios.

Tal modalidade apresentou-se como alternativa à possível resistência dos pesquisadores à realização de mais uma atividade e como forma de convencimento do impacto gerado à produção e aos pesquisadores com o depósito em acesso aberto em um RD. O que se pensava era em implantar o repositório com o depósito mediado e posteriormente avançar para o autoarquivamento.

Contudo, o software possibilita o depósito utilizando importação de dados, ferramenta esta que passou a ser utilizada de forma inovadora por alguns repositórios brasileiros, a exemplo do Repositório Institucional da Universidade Estadual Paulista (Unesp), motivado pela recomendação do depósito de artigos por parte de uma agência de fomento, a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Nesse sentido, Vidotti et al. (2016) afirmam que a decisão de povoar o repositório utilizando coleta automática, a partir do reuso de metadados de diferentes fontes, embasou-se na problemática de tempo necessário para a alimentação por autoarquivamento, bem como do tempo necessário pelo tratamento dos dados oriundos desta modalidade.

Esta modalidade de depósito, tem sido chamada de coleta automática, e utiliza ferramentas de interoperabilidade de dados e outros artefatos tecnológicos para a coleta dos metadados e dos arquivos diretamente das bases de dados, nas quais há grande quantidade de periódicos e registros. É possível ainda a coleta de registros ou integração com outros sistemas disponíveis na própria instituição. Realizada a coleta, os metadados são tratados tendo em vista a política adotada institucionalmente e então são importados para o repositório.

Muito embora a coleta automática ainda seja pouco referendada como forma de depósito pela literatura, não resta dúvida de que a integração dos sistemas disponíveis e o reuso de metadados é a alternativa mais viável para a alimentação e o povoamento dos repositórios.

Nesse sentido, a Budapest Open Access Initiative (2012) também estabelece que “os repositórios AA devem possuir meios para e permitir a colheita dos artigos e re-depósito em outros repositórios AA”.

Nada impede, contudo, que o RD utilize mais que uma forma de povoamento, é importante que a política explicita a(s) forma(s) de entrada dos dados, fato este que determinará a velocidade do crescimento do repositório.

## FLUXO DE TRABALHO

Para estabelecer o fluxo de trabalho, é imprescindível ter clareza das formas de povoamento do repositório, lembrando que este fluxo pode ser definido individualmente por coleção.

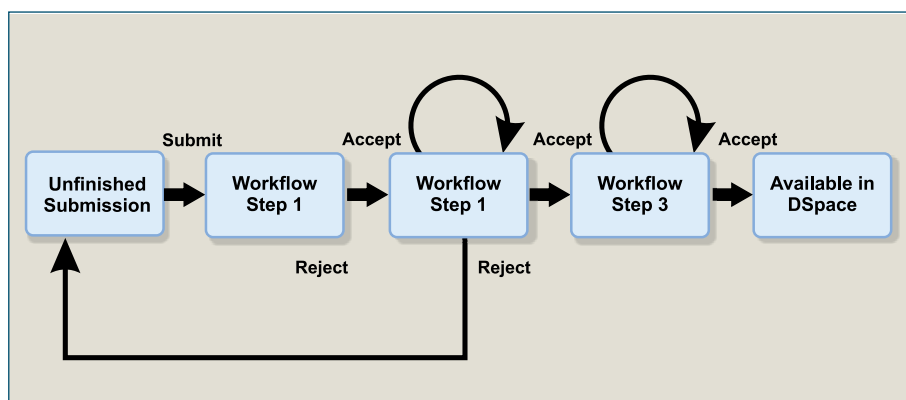


Figura 2 – Workflow steps

Fonte: Kainerstorfer e Perkins (2009).

O software DSpace possibilita que um item transite em diferentes estágios de um fluxo de atividades, que inclui a submissão, a verificação e edição dos metadados e posterior disponibilização, conforme parametrizado pelo administrador da coleção (Figura 2). Esta funcionalidade permite que uma submissão possa passar por até três etapas até sua disponibilização on-line. O processo permite assegurar a adequação do conteúdo às políticas do repositório e da coleção, a qualidade do arquivo, do conteúdo, dos metadados, a adequação da licença de direito autoral e formas de acesso permitidas.

O processo de submissão, revisão e disponibilização pode ser utilizado no caso de autoarquivamento e depósito mediado. No caso de coleta automática, ademais da possibilidade de editar o item na própria plataforma, pode-se fazê-lo antes da entrada dos dados na base.

Independentemente da forma de entrada, cumpre considerar que há a necessidade do tratamento adequado dos itens para que a visibilidade seja assegurada.

Na política, o fluxo de trabalho precisa ser registrado de forma clara, atribuindo responsabilidades a cada um dos atores envolvidos no processo e, quando possível, elencando prazos para a conclusão de cada uma das etapas.

## TRATAMENTO DA INFORMAÇÃO

Há alguns cuidados que devem ser tomados na disponibilização de um item. A política precisa estabelecer claramente como será realizado o tratamento da informação, etapa esta realizada ou supervisionada por profissionais da informação.

Vale destacar que é durante o tratamento da informação que o item é analisado e descrito e, considerando o complexo cenário de quantidade de documentos disponíveis na web, depende dele a recuperação.

Atendendo ao fluxo de trabalho parametrizado para a coleção, o item pode ser depositado e passar por um ou mais estágios de revisão de metadados ou ser imediatamente disponibilizado no RD e acessível na internet.

A forma de depósito também influenciará este processo. Em geral, quando do autoarquivamento, há etapas de revisão; no caso do depósito mediado, pode haver etapa de revisão; e, no caso do depósito automático, o tratamento da informação pode ocorrer antes da importação dos dados ou enquanto o item ainda não está visível para o usuário final. Recomenda-se que o tratamento da informação seja feito por profissional capacitado.

O perfil de aplicação terá determinado os campos a serem preenchidos na descrição, e o arquivo digital deverá estar no formato especificado e pronto para a disponibilização. É importante que o arquivo seja nomeado com o título do documento para favorecer a sua visibilidade web.

No tratamento, deverá ser controlada a qualidade da descrição do conteúdo utilizando-se os metadados estabelecidos para a tipologia pelo perfil de aplicação. Há ainda a necessidade de verificar a correspondência entre eles e o arquivo carregado.

Além disso, há de se considerar a viabilidade da padronização de autoridades para que os dados tenham maior relevância no momento da recuperação da informação por parte dos mecanismos de busca.

Cabe ainda ao tratamento da informação a verificação das condições de disponibilização e aplicação de licenças, quando presentes.

Ainda a critério do comitê gestor, podem ser aplicadas outras padronizações, como o uso de um código de classificação.

## PRESERVAÇÃO DIGITAL

Para Ferreira (2006), preservação digital é o conjunto de atividades e processos utilizados para assegurar o acesso contínuo e em longo prazo à informação digital, garantindo que permaneça acessível, com qualidade e autenticidade para a interpretação futura, utilizando-se de uma plataforma tecnológica.

Arellano e Oliveira (2016) apresentam como soluções parciais a “migração, emulação, arqueologia digital, criptografia, metadados, formatos-padrão e software livre”.

Ferreira (2006) menciona que uma das primeiras estratégias de preservação é a tecnológica, na qual o contexto utilizado na concepção do objeto é preservado. Nesse sentido, elenca: refrescamento, emulação, migração/conversão, migração para suportes analógicos, atualização de versões, conversão para formatos concorrentes, normalização, migração a pedido, migração distribuída, encapsulamento.

Vale lembrar que apenas a preservação digital poderá assegurar que o documento possa ser utilizado em longo prazo. Para tanto, é relevante atentar que se trata não apenas de arquivos, mas também de metadados (descritivos, de direito autoral e técnicos) e a estrutura do documento.

No momento da definição da política, é necessário estabelecer a forma de atuação do RD no que concerne à preservação digital, atentando-se às normas e legislação vigentes.

## ATUALIZAÇÃO

Independente do software a ser utilizado, o funcionamento adequado depende de um conjunto de sistemas interligados que levam ao seu funcionamento adequado: base de dados, interface, comunicação com servidor, indexação e estatísticas.

É importante lembrar que todos estes elementos são interdependentes e geram impactos significativos no momento das atualizações.

É salutar definir uma política de atualização destes sistemas, de forma que o funcionamento do repositório não seja ameaçado. Além disso, o próprio software utilizado deve ser atualizado sempre que uma nova versão estável seja disponibilizada.

Nesse sentido, recomenda-se manter uma instância de testes na qual a atualização da versão possa ser verificada, antes de aplicar na máquina de produção.

## QUADRO SINTÉTICO

Ao estabelecer	Lembre-se de
Objetivo	Alinhar o RD às políticas locais já existentes; definir os objetivos da política, do RD e sua área de atuação.
Equipe, instâncias e atribuições	Estabelecer as instâncias envolvidas; definir atribuições; mapear competências necessárias para a formação da equipe.
Mandato e mecanismos de acompanhamento	Definir um mandato de depósito obrigatório; requerer o depósito imediato ou no menor tempo possível; apresentar as formas de acompanhamento do depósito.
Arquitetura da informação	Planejar a arquitetura de informação.
Gestão de coleções	Definir a estrutura; evitar a duplicidade de rótulos; estabelecer o fluxo de trabalho.
Metadados	Desenvolver o perfil de aplicação; utilizar esquemas de metadados normalizados; apresentar os requisitos de descrição para cada elemento.
Tipologias, formatos e tamanho de arquivos	Elencar as tipologias aceitas; definir os formatos de arquivos aceitos; avaliar constantemente a capacidade do servidor que hospeda o serviço.
Direito autoral e embargo	Estabelecer diretrizes para incentivar a publicação em OA; definir as formas de acesso aos materiais depositados, em função do direito autoral: acesso aberto, acesso embargado, acesso restrito.
Formas de povoamento	Autoarquivamento; depósito mediado; coleta automática.
Fluxo de trabalho	Definir o workflow; indicar as responsabilidades em cada etapa do fluxo. Nota: Esta definição é diretamente afetada pela forma de povoamento.
Tratamento da informação	Estabelecer tratamento adequado da informação, em sintaxe e semântica; manter qualidade dos metadados; padronizar autoridades; disponibilizar obras de acordo com a licença, atentando-se ao direito autoral.
Preservação digital	Utilizar de normas e padrões de preservação de longo prazo.
Atualização	Manter os softwares atualizados. Nota: Caso utilize software livre, quando do desenvolvimento do código, reportar aos desenvolvedores.

Quadro 1 – Elementos a serem considerados na elaboração de política para RD

Fonte: Autoria própria (2017).

Considerando a possibilidade de utilização de software livre, recomenda-se que todas as alterações no código fonte sejam adequadamente planejadas, evitando problemas futuros de atualização, que poderiam deixar a instalação marginalizada em relação ao software. Para as alterações realizadas, uma boa prática é reportar à comunidade de desenvolvedores, com vistas à possibilidade de incorporação em versões futuras, contribuindo, desta maneira, com o desenvolvimento do sistema.

## CONSIDERAÇÕES FINAIS

O panorama do OA no Brasil ainda é de desenvolvimento. Embora muitas instituições tenham obtido sucesso na implantação e disponibilização de repositórios digitais, ainda não se pode afirmar que esta é uma ferramenta usual.

Ainda há um caminho a percorrer no convencimento e na motivação de pesquisadores, instituições e agências de fomento, em prol de práticas e ferramentas de OA. A esse respeito a Budapest Open Access Initiative (2012) reforça que:

Devemos fazer mais para consciencializar os editores de revistas, os editores científicos, revisores e pesquisadores dos padrões de conduta profissional para publicação em AA, por exemplo, quanto ao licenciamento, processo editorial, apelo à submissão de artigos em AA, identificação de propriedade e gestão de taxas de publicação. Os editores científicos, revisores e pesquisadores devem avaliar as oportunidades de envolvimento com editores de revistas na base destes padrões de conduta profissional. Quando os editores não agirem de acordo com esses padrões devemos ajudá-los a melhorar, como um primeiro passo.

Alterar este cenário requer envolvimento e esforço e estará embasado num conjunto de práticas e documentos norteadores. Nesse sentido, considerando a gestão de repositórios digitais, a definição de políticas claras favorecerá o desenvolvimento de forma a atingir resultados positivos.

É necessário formalizar a existência do repositório, bem como estabelecer seus objetivos e formas de atuação, de forma a torná-lo uma fonte de informação digital confiável e permanente, cujo crescimento se dará de forma ordenada e embasada, gerando benefícios inegáveis, não apenas às instituições e autores envolvidos, mas de forma mais ampla, à sociedade.

Nesse sentido, é fundamental compreender a relevância da definição de políticas que possam, além de nortear o desenvolvimento do serviço, assegurar sua permanência.



## REFERÊNCIAS

- ARELLANO, M. A. M.; OLIVEIRA, A. F. de. Gestão de repositórios de preservação digital. **Rev. Digit. Bibliotecon. Cienc. Inf.**, v. 14, n. 3, p. 465-483, set./dez. 2016. Disponível em: <<http://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8646346/pdf>>. Acesso em: 05 maio 2017.
- BUDAPEST OPEN ACCESS INITIATIVE. **Dez anos da iniciativa de Budapeste em acesso aberto**: a abertura como caminho a seguir. 2012. Disponível em: <<http://www.budapestopenaccessinitiative.org/boai-10-translations/portuguese-brazilian-translation>>. Acesso em: 13 abr. 2017.
- BUDAPEST OPEN ACCESS INITIATIVE. **Iniciativa de Budapeste pelo acesso aberto**. 2002. Disponível em: <<http://www.budapestopenaccessinitiative.org/translations/portuguese-translation>>. Acesso em: 13 abr. 2017.
- CAMARGO, L. S. de A. de; VIDOTTI, S. A. B. G. **Arquitetura da informação**: uma abordagem prática para o tratamento de conteúdo e interface em ambientes informacionais digitais. Rio de Janeiro: LTC, 2011.
- CASSARRO, A. C. **Sistemas de informação para tomada de decisões**. São Paulo: Pioneira, 1999.
- CORRADO, E. M.; JAFFE, R. Transforming and enhancing metadata for enduser discovery: a case study. **JLIS.it.**, v. 5, n. 2, jul. 2014. Disponível em: <<https://www.jlis.it/article/view/10069/9352>>. Acesso em: 05 maio 2017.
- DAVENPORT, T. H. **Ecologia da informação**: porque só a tecnologia não basta para o sucesso na era da informação. São Paulo: Futura, 1998.
- FERREIRA, M. **Introdução à preservação digital**: conceitos, estratégias e actuais consensus. Guimarães, PT: Escola de Engenharia da Universidade do Minho, 2006. Disponível em: <<https://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>>. Acesso em: 05 fev. 2017.
- JARDIM, J. M.; SILVA, C. A.; NHARRELUGA, R. S. Análise de políticas públicas: uma abordagem em direção às políticas públicas de informação. **Perspectivas em Ciência da Informação**. Belo Horizonte, v. 14, n. 1, p. 2-22, 2009. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/743/535>>. Acesso em: 13 abr. 2017.
- KAINERSTORFER, C.; PERKINS, H. **Intro to DSpace**. 2009. Disponível em: <[http://www.tdl.org/wp-content/uploads/2009/08/IntroToDSpace\\_081920091.pdf](http://www.tdl.org/wp-content/uploads/2009/08/IntroToDSpace_081920091.pdf)>. Acesso em: 15 abr. 2017.
- LEITE, F. C. L. **Como gerenciar e ampliar a visibilidade da produção científica brasileira**: repositórios institucionais de acesso aberto. Brasília: Ibict, 2009. Disponível em: <<http://livroaberto.ibict.br/bitstream/1/775/4/Como%20gerenciar%20e%20ampliar%20a%20visibilidade%20da%20informa%C3%A7%C3%A3o%20cient%C3%ADfica%20brasileira.pdf>>. Acesso em: 13 abr. 2017.
- LEITE, F. C. L. et al. **Boas práticas para a construção de repositórios institucionais da produção científica**. Brasília: Ibict, 2012. Disponível em: <<http://livroaberto.ibict.br/bitstream/1/703/1/Boas%20pr%C3%A1ticas%20para%20a%20constru%C3%A7%C3%A3o%20de%20reposit%C3%B3rios%20institucionais%20da%20produ%C3%A7%C3%A3o%20cient%C3%ADfica.pdf>>. Acesso em: 13 abr. 2017.

MCGEE, J. V.; PRUSAK, L. **Gerenciamento estratégico da informação**. Rio de Janeiro: Campus, 1994.

MORAES, G. D. de A.; ESCRIVÃO FILHO, E. A gestão da informação diante das especificidades das pequenas empresas. **Ciência da Informação**, Brasília, v. 35, n. 3, p. 124-132, dez. 2006. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19652006000300012&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652006000300012&lng=en&nrm=iso)>. Acesso em: 13 abr. 2017.

MORVILLE, P.; ROSENFELD, L. **Information architecture for the World Wide Web**. Sebastopol, CA: O'Reilly, 2006.

RODRIGUES, E. **Kit de políticas open access**. 2009. Disponível em: <<http://projeto.rcaap.pt/index.php/lang-pt/consultar-recursos-de-apoio/remository?func=startdown&id=97>>. Acesso em: 14 abr. 2017.

SILVA, T. E. da. Política de informação na pós-modernidade: reflexões sobre o caso do Brasil. **Revista Informação & Sociedade: Estudos**, João Pessoa, v. 1, n. 1, 1991. Disponível em: <[http://basessibi.c3sl.ufpr.br/brapci/\\_repositorio/2010/12/pdf\\_55cc575a2c\\_0013996.pdf](http://basessibi.c3sl.ufpr.br/brapci/_repositorio/2010/12/pdf_55cc575a2c_0013996.pdf)>. Acesso em: 13 abr. 2017.

STAIR, R. M. **Princípios de sistemas de informação**: uma abordagem gerencial. Rio de Janeiro: LTC Livros Técnicos e Científicos, 1998.

TOMAÉL, M. I.; SILVA, T. E. da. Repositórios institucionais: diretrizes para políticas de informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. **Anais eletrônicos...** Salvador: EDUFBA, 2007. Disponível em: <<http://www.enancib.ppgci.ufba.br/artigos/GT5--142.pdf>>. Acesso em: 05 nov. 2016.

VIDOTTI, S. A. B. G. et al. Coleta automática para o povoamento de repositórios digitais: conversão de registros utilizando XSLT. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., 2016, Salvador. **Anais eletrônicos...** Salvador: UFBA, 2016. Disponível em: <<http://www.ufpb.br/evento/lti/ocs/index.php/enancib2016/enancib2016/paper/viewFile/4092/2552>>. Acesso em: 05 fev. 2017.

WURMAN, R. S. **Information architects**. Zurich: Graphis, 1996.

# REPOSITÓRIO INSTITUCIONAL UNESP: INOVAÇÃO NO PROCESSO DE CRIAÇÃO

Silvana Aparecida Borsetti Gregorio Vidotti  
Flávia Maria Bastos  
Felipe Augusto Arakaki





## BACKGROUND: A EVOLUÇÃO DO REPOSITÓRIO INSTITUCIONAL UNESP

Visando à ampliação da divulgação das pesquisas financiadas no Estado de São Paulo, a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), no final do ano de 2012, exigiu que as três universidades estaduais paulistas (Universidade Estadual Paulista Júlio de Mesquita Filho – UNESP, Universidade de São Paulo – USP, Universidade Estadual de Campinas – UNICAMP) construíssem repositórios institucionais com o objetivo de divulgar os trabalhos financiados por ela.

Naquele momento, a UNESP não possuía um repositório de produção bibliográfica, mas utilizava duas bases para disponibilização de sua produção acadêmica, a C@thedra, que armazenava e disponibilizava teses e dissertações, e a Biblioteca Digital C@pelo, que disponibiliza os trabalhos de conclusão de curso da graduação. Entretanto, não existia um ambiente para disponibilização de artigos científicos e trabalhos em eventos, entre outros tipos de produções bibliográficas.

Nesse contexto, os trabalhos para construção do Repositório Institucional UNESP (RI UNESP) iniciaram em 2013, com a Portaria UNESP n. 88, que instituiu o Grupo Gestor da Política do Repositório Institucional UNESP (UNIVERSIDADE ESTADUAL PAULISTA, 2013). Junto com o repositório da USP e da UNICAMP, o RI UNESP iria compor o Repositório de Produção Científica do Conselho de Reitores das Universidades Estaduais Paulistas (CRUESP) (ASSUMPÇÃO et al., 2014; VIDOTTI et al., 2015a; VIDOTTI et al., 2016).

Todavia, para a construção do RI UNESP, foi necessário planejar e estruturar toda a base de dados em padrões internacionalmente reconhecidos. Nesse contexto, a fase de planejamento configurou-se em atender primeiramente a demanda solicitada pela FAPESP com a disponibilização dos artigos dos autores da UNESP publicados nas bases *Scientific Electronic Library Online* (SciELO) e *Web of Science* (WoS).

Uma vez definidas as metas, o próximo passo foi a definição dos metadados que iriam descrever os objetos digitais que seriam disponibilizados no RI UNESP. Nesse contexto, a estrutura proposta pelo padrão *Dublin Core* (DC) correspondia

às expectativas para a descrição dos objetos digitais do RI UNESP. Entre as características do *Dublin Core* e de acordo com Baptista e Machado (2001), estão:

- a) **simplicidade**: elementos de fácil entendimento, que possibilita a não especialistas descrever um recurso na Web;
- b) **interoperabilidade semântica**: os metadados definidos possuem escopo de uso genérico abarcando diversos metadados de outros padrões, possibilitando, assim, a interoperabilidade;
- c) **consenso internacional**: reconhecimento e aceitação Internacional do *Dublin Core*. Atualmente, as normas regularizam os quinze metadados para descrição de recursos. Entre elas estão as normas da *The Internet Engineering Task Force* (IETF) – RFC 5013 de 2007, da *American National Standards Institute* (ANSI)/*National Information Standards Organization* (NISO), Z39.85:2012 de 2012 e da *International Organization for Standardization* (ISO) 15836-1:2017;
- d) **extensibilidade**: o *Dublin Core* possibilita acrescentar novos metadados que não estão previstos em seu escopo;
- e) **modularidade de metadados na web**: o padrão permite utilizar outros padrões de metadados para compor a descrição dos documentos.

Para promover a interoperabilidade do RI UNESP com outras bases, sem perder as características do *Dublin Core* como, por exemplo, a possibilidade de extensão dos metadados para atender as necessidades descritivas do RI UNESP, foi considerada a construção de um Perfil de Aplicação *Dublin Core* (*Dublin Core Application Profile* – DCAP). Segundo Nilsson, Baker e Johnston (2008, tradução nossa), “Um AP [Perfil de Aplicação] é um processo que contribui no planejamento de definições para construção de: requisitos funcionais, modelo de domínio, conjunto de descrição, guia de uso e guia de sintaxe”.

Ainda segundo Coyle e Baker (2009, tradução nossa), “Um DCAP é um documento (ou conjunto de documentos) que especifica e descreve os metadados utilizados para uma aplicação particular”. Para os autores, o DCAP:

- a) descreve o que a comunidade quer realizar com a sua aplicação (requisitos funcionais);
- b) caracteriza os tipos de objetos/recursos descritos pelos metadados e suas relações (Modelo de domínio);
- c) enumera os termos de metadados a serem utilizados e as regras para a sua utilização (Descrição conjunto de perfis e Diretrizes de uso);
- d) define a sintaxe máquina que será utilizada para codificar os dados (Diretrizes de sintaxe e formatos de dados) (COYLE; BAKER, 2009, tradução nossa).

No Quadro 1, é apresentada uma síntese do Perfil de Aplicação RI UNESP.

Metadados	Termos utilizados	Obrig.	Repet.	Doc.	Idio.
Orientador	dc.contributor.advisor	A	R	DM, TD, TCC	N
Autor	dc.contributor.author	O	R	G	N
Instituição	dc.contributor.institution	O	R	G	N
Data de publicação	dc.date.issued	O	NR	G	N
Nota	dc.description	F	R	G	N
Resumo	dc.description.abstract	O	R	G	S
Afiliação	dc.description.affiliation	A	R	ART	N
Afiliação do autor UNESP	dc.description.affiliationUnesp	A	R	ART	N
Agência de fomento/ financiamento	dc.description.sponsorship	A	R	G	N
Número do processo/ financiamento	dc.description.sponsorshipId	A	R	G	N
Extensão do recurso	dc.format.extent	F	NR	G	N
URL do recurso informacional	dc.identifier	A	R	G	N
Identificador do Aleph (SYS)	dc.identifier.aleph	A	NR	G	N
Identificador da CAPES para Programas de Pós-Graduação	dc.identifier.capes	A	R	G	N
Referência do documento	dc.identifier.citation	A	NR	G	N
DOI	dc.identifier.doi	A	R	G	N
Nome do arquivo ou o caminho do arquivo	dc.identifier.file	A	R	G	N
ISSN	dc.identifier.issn	A	R	ART	N
Identificador do Currículo Lattes	dc.identifier.lattes	O	R	ART	N
Identificador PubMed	dc.identifier.pubmed	A	NR	ART	N
Identificador PubMed Central	dc.identifier.pmc	A	NR	ART	N
Identificador SciELO	dc.identifier.scielo	A	NR	ART	N
Identificador Scopus	dc.identifier.scopus	A	R	ART	N



Metadados	Termos utilizados	Obrig.	Repet.	Doc.	Idio.
Identificador do registro no Repositório	dc.identifier.uri	O	NR	G	N
Identificador Web of Science	dc.identifier.wos	A	R	ART	N
Idioma	dc.language.iso	O	R	G	N
Publicador	dc.publisher	A	R	G	N
Fonte	dc.relation.ispartof	A	NR	ART	N
Fonte do registro	dc.source	A	NR	G	N
Palavra-chave	dc.subject	O	R	G	S
Título	dc.title	O	NR	G	S
Título alternativo	dc.title.alternative	A	R	G	S
Tipo	dc.type	O	NR	G	N
Condições de acesso	dc.rights.accessRights*	O	NR	G	N
Licença de uso	dcterms.license	A	R	G	N
Detentor dos direitos	dcterms.rightsHolder	A	R	G	N
Unidade	unesp.campus	O	R	G	PT
Departamento	unesp.department	A	R	G	PT
Programa de Pós-Graduação	unesp.graduateProgram	A	R	G	PT
Área de concentração	unesp.knowledgeArea	A	R	G	PT
Linha de pesquisa	unesp.researchArea	A	R	G	PT
Data para a disponibilização do conteúdo integral	unesp.embargo	A	NR	G	PT

Quadro 1 – Perfil de Aplicação RI UNESP

Fonte: Vidotti et al. (2016, p. 12-13).

Legenda: O - Obrigatório; A - Obrigatório se aplicável ao documento que está sendo descrito; F - Facultativo; R - Repetível; NR - Não repetível; G - Todos os documentos; ART- Artigos (inclui: resenhas, editoriais, cartas, notas, trabalhos e resumos apresentados em eventos); DM - Dissertação de mestrado; TD - Tese de doutorado; TL - Tese de livre-docência; TCC - Trabalho de conclusão de curso; L - Livros; S - Sim, indicar o idioma; N - Não indicar o idioma; PT - Indicar sempre o idioma 'pt' (português).

Pensando em tornar disponível os conteúdos gerados internamente na Unesp, outras demandas foram solicitadas para inclusão no RI UNESP. A exemplo disso foram incluídos as imagens das capas e o Podcast relacionados às publicações dos livros da Editora UNESP – selo Cultura Acadêmica e, em comemoração



aos 40 anos da UNESP, foram incluídos a partitura e o hino da UNESP. Além disso, foi realizado um estudo de inclusão de documentos administrativos tais como as portarias e as resoluções da Universidade. A cada novo tipo de recurso a ser incorporado são estudadas a estrutura e as necessidades descritivas necessárias à melhor forma de recuperação pelos usuários, o que implica em ampliação do perfil de aplicação RI UNESP.

## METODOLOGIA DE COLETA DE DADOS RI UNESP

Diante da demanda de povoamento do repositório em um curto espaço de tempo, a UNESP decidiu inovar na inclusão de registros a partir da coleta automática de metadados, ao invés do autoarquivamento comumente realizado em repositórios. Segundo Vidotti et al. (2016, p. 3):

A decisão de incluir objetos a partir de coletas automáticas de registros de diferentes fontes foi tomada, pois, a implementação do autoarquivamento [...] poderia demandar muito tempo, uma vez que seria necessário criar campanhas de divulgação junto à comunidade, e ainda não seria garantido que o depósito fosse realizado pelo pesquisador. Outro possível problema decorrente do autoarquivamento seria o tempo que demandaria para o controle de qualidade das informações fornecidas pelos pesquisadores e a verificação dos direitos autorais dos documentos.

A metodologia inovadora desenvolvida pelo RI UNESP consiste no reuso de dados de diferentes bases de dados internas e externas à UNESP para povoar/alimentar o repositório. Conforme relatado por Assumpção et al. (2014) e Vidotti et al. (2015b), o processo de conversão de registros foi, inicialmente, realizado a partir das bases *Web of Science*, SciELO e Scopus. Adicionalmente, foram coletadas as produções bibliográficas a partir dos Currículos Lattes dos docentes da UNESP (VIDOTTI et al., 2015a), do PubMed e de bases internas como a do catálogo da rede de bibliotecas e do sistema de Recursos Humanos da Unesp.

A metodologia utilizada para conversão dos registros tem como base as folhas de estilo *eXtensible Stylesheet Language Transformation* (XSLT) e é composta por 9 passos, conforme assinalam Vidotti et al. (2016).

### PRIMEIRO PASSO: COLETA DOS REGISTROS EM XML (VIA INTERFACE OU EM LOTE)

A busca dos registros de autores da Unesp inicia-se pela afiliação do pesquisador. Apesar de existir uma portaria que regulamenta a forma de padronização da afiliação, muitos registros aparecem com formas variadas, assim, as buscas

acabam sendo mais demoradas, pois devem ser realizadas buscas para cada afiliação, por exemplo: Univ. Estadual Paulista; UNESP; Universidade Estadual Paulista; UNESP – Univ Estadual Paulista; São Paulo *State University*, entre outras variações possíveis.

Para coletas realizadas na base da *Web of Science*, é utilizado o web service, que permite consultar e baixar os registros de forma automática. Inicialmente, os registros em XML da *Scopus* foram adquiridos mediante compra, na época, era a única forma de obter esses registros; hoje, porém, os registros são obtidos de forma automática. Seguindo a coleta de outras bases de dados, os registros da base *PubMed* são coletados, via interface da própria base, pelos bibliotecários do Repositório. Já os registros da *SciELO* são coletados por um programa desenvolvido especificamente para essa atividade. Os registros do *Currículo Lattes* são coletados a partir da seleção dos identificadores Lattes dos docentes da Universidade e baixados por um programa também desenvolvido pela equipe do RI UNESP (VIDOTTI et al., 2015b; VIDOTTI et al., 2016).

Os registros de teses, dissertações e trabalhos de conclusão de curso de graduação, armazenados nas bases C@thetra e C@pelo, são coletados a partir do software Aleph, utilizado pela Rede de Bibliotecas da UNESP. Os dados são convertidos de MARC 21/ISO 7909 para MARC 21/MARCXML a partir do software *MarcEdit*, depois é realizada mais uma conversão de MARCXML para *Dublin Core/XML*, gerando, finalmente, o formato de saída dos dados (VIDOTTI et al., 2016).

Ainda, são coletados registros bibliográficos em plataformas como *Open Journal Systems* (OJS), *Open Conference Systems* (OCS) e DSpace, nas quais os gestores tenham habilitado o protocolo *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH), o que torna o processo de coleta de metadados mais simplificado. O OAI-PMH é um mecanismo para coleta e interoperabilidade de dados e está disponível em diversas plataformas como OJS, OCS, DSpace, entre outras. Na Figura 1, tem-se um exemplo de como as informações são dispostas para coleta de dados.

Para extrair esses dados é utilizado o programa *MarcEdit*, no botão *Harvest OAI Records*. Na informação *Server* é inserido o endereço do servidor que está armazenando os dados do sistema. O elemento *Set Name* é opcional e pode ser utilizado quando se quer coletar apenas os dados de uma coleção específica, ou um único registro, e não se quer baixar todas as informações de uma determinada base de dados. No elemento *Metadata*, pode-se escolher algumas opções oferecidas pelo programa, como *Dublin Core*. Para salvar os dados, clique em *Harvest Raw Data* e em *Save folder* e escolha o local que deseja salvar, conforme exemplifica a Figura 2.

<https://repositorio.unesp.br/oai/request?verb=Identify>

DSpace OAI-PMH Data Provider      Identify   Sets   Records   Identifiers   Metadata Formats

Response Date: 2017-09-30 18:15:07

### Repository Information

Repository Name	Repositório Institucional UNESP
E-Mail Contact	repositoriounesp@reitoria.unesp.br
Repository Identifier	repositorio.unesp.br
Sample Identifier	oai:repositorio.unesp.br:11449/1234
Protocol Version	2.0
Earliest Registered Date	2013-09-30 17:40:58
Date Granularity	YYYY-MM-DD hh:mm:ss
Deletion Mode	transient

Figura 1 – OAI-PMH do RI UNESP  
Fonte: Repositório Institucional UNESP (2017).

## Harvesting: MarcEdit

Metadata Harvester

**Harvest OAI Data**

Server:  Link do OAI

Set Name:  Endereço da comunidade (caso queira os metadados de uma comunidade específica)

Metadata:  Formato de saída

Save Folder:  Apenas 1 registro

**Advanced Settings**

GetRecord:  Clique aqui para salvar no computador

ResumptionToken:

Start:  End:

☐ Translate to MARC-8

Timeout:  secs.

☒ Harvest Raw Data (save OAI data to local file system)

OK   Save   Close

Figura 2 – Ferramenta MarcEdit para coleta de dados  
Fonte: Adaptado de Reese (2016).

Após o *download*, o arquivo está pronto para ser trabalhado e ajustado para o Perfil de Aplicação RI UNESP. Nesse caso, como o RI UNESP utiliza como base o *Dublin Core*, o processo de mapeamento dos dados é mais rápido.

## SEGUNDO PASSO: ESTUDO DA ESTRUTURA DO XML

Ao analisar os registros advindos das bases de dados, observou-se que cada base utiliza uma estrutura própria, e não padrões de metadados internacionalmente reconhecidos. Isso dificulta o processo de coleta e conversão de dados, pois demanda um tempo da equipe para entender a estrutura inicial e projetar o mapeamento dos dados para o Perfil de Aplicação RI UNESP. Nessa etapa, conhecimentos da linguagem *eXtensible Markup Language* (XML) são fundamentais para entender a estrutura do registro.

Conforme o exemplo da Figura 3, para descrever o título do recurso, os registros da *Web of Science* utilizam a marcação *title type="item"* que expressa exatamente a informação desejada para inserir no metadado *dc.title* do Perfil de Aplicação RI UNESP. Na Scopus é utilizada a marcação *titletext* que está dentro da marcação *citation-title*. Nos registros da SciELO é utilizada a marcação *article-title* que está dentro da marcação *title-group*. Já nos registros da PubMed é utilizada a marcação *ArticleTitle*, enquanto que nos registros do Currículo Lattes é utilizada a marcação *DADOS-BASICOS-DO-ARTIGO TITULO-DO-ARTIGO*.

Plataformas como OJS e OCS disponibilizam e operam com o protocolo OAI-PMH, o que facilita a conversão dos registros bibliográficos, pois os dados podem ser extraídos na estrutura do *Dublin Core*, bastando separar as informações que serão incorporadas no RI UNESP em consonância com o metadado a ser incluído.

Web of Science	<pre>&lt;titles count="6"&gt;&lt;title type="source"&gt;CURRENT APPLIED PHYSICS&lt;/title&gt;&lt;title type="source_abbrev"&gt;CURR APPL PHYS&lt;/title&gt;&lt;title type="abbrev_iso"&gt;Curr. Appl. Phys.&lt;/title&gt;&lt;title type="abbrev_11"&gt;CURR APPL P&lt;/title&gt;&lt;title type="abbrev_25"&gt;CURR APPL PHYS&lt;/title&gt;&lt;title type="item"&gt;Structural refinement, optical and ferroelectric properties of microcrystalline Ba(Zr<sub>0.05</sub>Ti<sub>0.95</sub>)O<sub>3</sub> perovskite &lt;/title&gt;&lt;/titles&gt;</pre>
Scopus	<pre>&lt;citation-title&gt;&lt;titletext xml:lang="eng" original="y"&gt;Evaluation of alternative methods for the disinfection of toothbrushes&lt;/titletext&gt;&lt;/citation-title&gt;</pre>
SciELO	<pre>&lt;title-group&gt; &lt;article-title xml:lang="en"&gt;Nicotine-enhanced oxidation of low-density lipoprotein and its components by myeloperoxidase/H2O2/Cl- system&lt;/article-title&gt; &lt;/title-group&gt;</pre>
PubMed	<pre>&lt;ArticleTitle&gt;Long-Standing Motor and Sensory Recovery following Acute Fibrin Sealant Based Neonatal Sciatic Nerve Repair.&lt;/ArticleTitle&gt;</pre>
Lattes	<pre>&lt;DADOS-BASICOS-DO-ARTIGO NATUREZA="COMPLETO" TITULO-DO-ARTIGO="Estudo mineralógico do argilomineral paligorequita visando sua aplicação tecnológica"/&gt;</pre>

Figura 3 – Registros coletados em XML

Fonte: Vidotti et al. (2016, p. 11).

Nessa etapa, são criadas regras e padronizações para melhorar a qualidade das descrições, garantindo, assim, a integridade e a padronização dos dados inseridos no sistema. Para tanto, são utilizadas folhas de estilo baseadas em linguagem *eXtensible Stylesheet Language Transformation* (XSLT).

As folhas de estilo XSLT são documentos contendo conjuntos de regras escritas com a linguagem XSLT que, de modo geral, indicam ao software responsável pela transformação o que ele deve fazer com os dados de um arquivo XML para transformá-lo em outro arquivo (ASSUMPÇÃO et al., 2014, p. 6).

Como exemplo desse processo, no metadado `dc.contributor.institution` foram definidas algumas padronizações para melhorar a qualidade de suas informações e agilizar o processo de verificação, de modo que não é necessário ficar corrigindo registro por registro. As regras são aplicadas para a padronização, tanto do nome da Unesp, quanto de outras universidades, a exemplo da USP, UNICAMP e de Instituições de Ensino Superior.

O Quadro 2 apresenta parte da folha de estilo utilizada para conversão de registros da SciELO, bem como as regras para padronização de alguns metadados. A linha 1 representa uma anotação da regra do metadado (*dc.contributor.institution*). A linha 3 representa o local no qual a folha de estilo deverá buscar o registro (*article-meta/aff/institutio n*). A linha 4 representa o metadado para o qual a folha de estilo deverá enviar essas informações (*dcvalue element="contributor" qualifier="institution" language="pt"*), ou seja, elemento do Dublin Core – Contributor, com o qualificador *institution* e com uma codificação de valor no idioma português. A linha 5 (*xsl:choose*) descreve que poderão ocorrer múltiplas condições para o preenchimento desse metadado. A linha 7 indica outra anotação, informando que serão padronizados os nomes das principais instituições. A linha 9 define a primeira regra de transformação que indica, conforme consta na linha 10, que qualquer palavra minúscula definida entre as linhas 11 e 21 serão padronizadas. Na linha 22, é definido o valor que será preenchido, que no caso da UNESP, se o registro tiver na afiliação instituição algum termo relacionado entre as linhas 11 e 21, este será padronizado para Universidade Estadual Paulista (UNESP).

Além do mapeamento, a dificuldade na construção das folhas de estilo consiste em prever todas as formas variantes que podem existir nos registros. Nesse contexto, conforme são realizados os testes e criadas novas folhas de estilo, esses valores são atualizados para melhorar a qualidade dos registros a serem incluídos.

Esse é um processo utilizado para padronizar, por exemplo, o idioma, as principais agências de fomento, a construção da referência do artigo automaticamente e a organização da ordem do nome dos autores (sobrenome, nome).

```

1 <!-- dc.contributor.institution -->
2
3 <xsl:for-each select="article-meta/aff/institution">
4   <dcvalue element="contributor" qualifier="institution" language="pt">
5     <xsl:choose>
6
7       <!-- Padroniza os nomes das principais instituições -->
8
9       <xsl:when test="
10         functx:contains-any-of(lower-case(),
11         ('unesp',
12         'univ estadual paulista',
13         'universidade estadual paulista',
14         'paulista state univ',
15         'sao paulo state univ',
16         'são paulo state univ',
17         'state univ sao paulo',
18         'state univ são paulo',
19         'univ estad paulista',
20         'estadual paulista',
21         'mesquita filho'))">
22         <xsl:text>Universidade Estadual Paulista (UNESP)</xsl:text>
23       </xsl:when>

```

Quadro 2 – Parte da folha de estilo SciELO  
Fonte: Autoria própria (2013).

## TERCEIRO PASSO: CONVERSÃO DOS REGISTROS XML EM DSPACE-XML

A partir das regras estabelecidas, procede-se à conversão dos dados, que já saem padronizados no formato DC (Figura 4). Dessa forma, os dados são agrupados de acordo com o perfil de aplicação RI UNESP e torna-se necessária a inclusão de informações, conforme apontado na etapa anterior.

Uma vez adaptados ao padrão da Unesp, os dados deverão ser convertidos para um formato que o DSpace suporte, como o *Comma-Separated Values* (CSV). Optou em dividir as folhas de estilo em duas partes: a primeira consiste no mapeamento dos registros para o padrão *Dublin Core* (RI UNESP); a segunda, na conversão dos registros em *Dublin Core* (RI UNESP) para CSV. Essa decisão foi tomada devido à complexidade de construção de folhas de estilo e maior facilidade de detecção de possíveis erros. Nesse sentido, uma vez ajustados os registros para o padrão da Unesp, a próxima fase fica facilitada, pois não é necessário recriar os mesmos passos para cada nova folha de estilo.

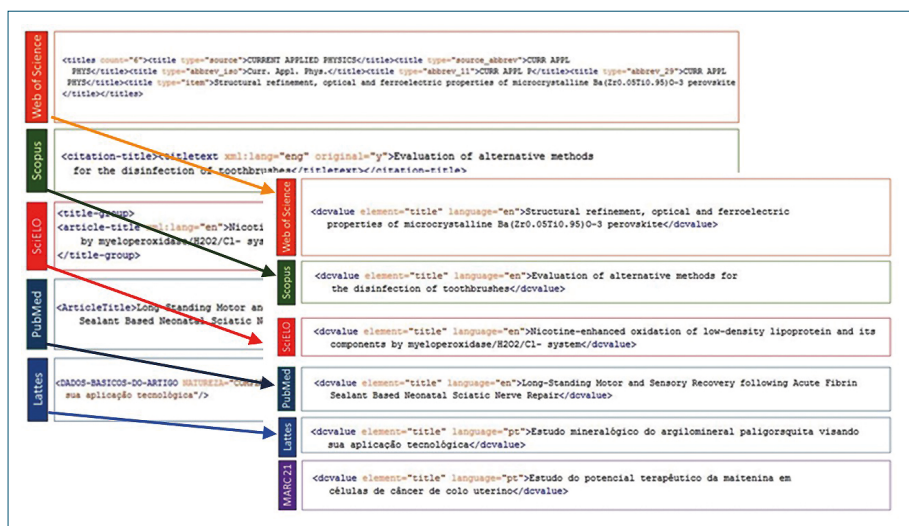


Figura 4 – Conversão de registros XML em DSpace-XML

Fonte: Autoria própria (2016).

## QUARTO PASSO: CONVERSÃO DOS REGISTROS DSPACE-XML PARA DSPACE-CSV

Após os registros estarem no padrão do RI UNESP, há uma transformação dos registros em XML para CSV. Para isso são criadas regras que direcionam as informações do XML para o CSV, conforme apresentado no Quadro 3, a seguir.

Os dados são convertidos conforme as instruções dadas pela folha de estilo e direcionadas da XML para CSV, conforme apresentado na Figura 5.

Ao final, no arquivo CSV cada coluna corresponde a um metadado e cada linha representa um registro bibliográfico (VIDOTTI et al., 2016).

```

1 <!-- dc.contributor.institution -->
2
3 <xsl:text>,&quot;</xsl:text>
4 <xsl:value-of select="replace(string-join(distinct-values(dcvalue[@element='contributor'
  and @qualifier='institution']),'|'), '&quot;,&quot;')" />
5 <xsl:text>&quot;</xsl:text>

```

Quadro 3 – Regras de conversão de XML para CSV

Fonte: Autoria própria (2013).



Scielo.xml x Scielo-DSpaceXML.xml x

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <records>
3   <record>
4     <dcvalue element="contributor" qualifier="author">Oliveira, Olga M.m.f.</dcvalue>
5     <dcvalue element="contributor" qualifier="author">Brunetti, Iguatemy L.</dcvalue>
6     <dcvalue element="contributor" qualifier="author">Khalil, Nafeh M.</dcvalue>
7     <dcvalue element="contributor" qualifier="institution">Universidade Estadual Paulista (UNESP)</dcvalue>
8     <dcvalue element="contributor" qualifier="institution">Universidade Estadual Paulista (UNESP)</dcvalue>
9     <dcvalue element="contributor" qualifier="institution">UNICENTRO Departamento de Farmácia</dcvalue>
10    <dcvalue element="date" qualifier="issued">2015-03-01</dcvalue>
  </record>
</records>

```

	A	B	C	D	E	F	G	H
1	id	collection	dc.contributor	dc.contributor	dc.contributor	dc.date.issued	dc.description	dc.description
2	+	11449/77075	Sarti, Flávia	Oliveira, Kar	Universidade	2013-11-16	The work aims to identify a	Est
3	+	11449/77075	Souza Neto, Recco, Keth	Universidade	2014-09-01	This study deals with teach	Est	
4	+	11449/11029	Brandão, Clá Kemper, Ber	Universidade	2015-02-13	The objective of this study is	O o	
5	+	11449/77162	Damasceno, Dallaqua, Bru	Universidade	2015-02-23	Natural products such as e	Pro	
6	+	11449/77090	Castiglia, Ya	Sudre, Eliana	Universidade	2015-02-13		
7	+	11449/77191	Fernandes Jo Andrade, Bru	Universidade	2015-02-23	Natural products such as e	Pro	
8	+	11449/77215	Oliveira, Clau Rossini, Bru	Universidade	2015-02-23			
9	+	11449/77090	Modolo, Norm Cabral, Luca	Universidade	2015-02-23			
10	+	11449/77222	Henry, Raoul Silva, Carolin	Universidade	2015-02-26			

Figura 5 – Conversão dos registros DSpace-XML para DSpace-CSV  
Fonte: Autoria própria (2016).

## QUINTO PASSO: ETAPA DE DE DUPLICAÇÃO DOS REGISTROS

Após a conversão dos registros para DSpace em CSV, é necessário verificar se alguns registros já foram inseridos no RI UNESP. Assim, primeiro verifica-se a existência de um mesmo *Digital Object Identifier* (DOI), e, se já houver o mesmo DOI cadastrado, o registro é excluído da planilha de entrada de dados. Em seguida, são reunidos todos os títulos dos trabalhos com o respectivo ano de publicação e esses são comparados com os registros do repositório. Com o decorrer do tempo, observou-se que, em algumas bases de dados, os títulos dos artigos estavam traduzidos em outro idioma, o que poderia duplicar registros no Repositório. Diante disso, começou-se a comparar títulos e subtítulos com seus respectivos anos.

## SEXTO PASSO: VERIFICAÇÃO MANUAL DE CADA REGISTRO

Na verificação dos dados, os registros foram verificados um a um com o objetivo de:

- identificar, a partir das informações de afiliação, se realmente eram parte da produção bibliográfica institucional; os artigos em que a UNESP não consta em nenhuma das afiliações são removidos;
- completar os dados que não puderam ser obtidos a partir dos dados coletados, por exemplo, as instituições de outros autores, as agências de fomento, o resumo, o título em outro idioma;
- corrigir possíveis erros, por exemplo, ordem dos autores e o endereço correto para o acesso online.



## SÉTIMO PASSO: COLETA DO PDF

A verificação das licenças consistiu em analisar as permissões de acesso (acesso aberto ou acesso restrito) e de arquivamento (permitido ou não o arquivamento da versão final em repositórios institucionais). Para essa verificação foi utilizado o serviço SHERPA/RoMEO de âmbito internacional, Diadorim para periódicos brasileiros e DULCINEA para periódicos espanhóis, além das políticas dos publicadores dos periódicos. Durante esse processo, observou-se que muitos periódicos não explicitam claramente sua política de direitos autorais, principalmente em relação ao depósito do artigo em repositórios institucionais.

## OITAVO PASSO: MAPEAMENTO DOS REGISTROS COM AS COLEÇÕES

A partir das informações de afiliação e de autoria, foram incluídos, no campo *collection* de cada registro, os códigos *Handle* das coleções às quais o objeto pertencia. Para registros que pertenciam a mais de uma coleção, os códigos das coleções eram separados dentro do campo *collection* por duas barras verticais (||).

Para realizar o mapeamento, é utilizado o metadado *dc.description.affiliationUnesp*, pois possui informações sobre departamento, programa de pós-graduação e unidade universitária a que o autor é vinculado, facilitando sua identificação e o mapeamento de sua produção bibliográfica. Quando não for possível identificar o Departamento de Ensino ou Programa de Pós-Graduação do autor por esse metadado, é realizado um mapeamento pelo metadado *dc.contributor.author*, a partir de um arquivo XML com o nome de cada Departamento de Ensino e do Programa de Pós-Graduação com sua respectiva lista de docentes. Os arquivos que não puderam ser identificados ficam em uma coleção para serem mapeados manualmente.

## NONO PASSO: IMPORTAÇÃO DOS REGISTROS NO REPOSITÓRIO

A importação dos registros pode ser realizada de duas formas. A primeira é via interface contendo os registros em CSV, limitando-se a 100 (cem) registros por arquivo. A segunda forma de importação é por linha de comando, e cada arquivo CSV pode conter até 1.000 (mil) registros, sendo esta última opção a mais rápida (VIDOTTI et al., 2016). Após importar todo o arquivo em CSV, são anexados, a partir de um programa desenvolvido pela equipe do RI UNESP, os arquivos em PDFs que podem ser disponibilizados no RI UNESP. Esse processo é feito pelo metadado *dc.identifier.file*, que é o mesmo nome do arquivo salvo, agilizando, assim, a inclusão dos arquivos nos registros.

## INTEGRAÇÃO E INTEROPERABILIDADE

A implantação do RI UNESP trouxe grandes benefícios para a visibilidade das pesquisas, dos pesquisadores e também para a Unesp. O uso de padrões internacionalmente reconhecidos garante maior integração e interoperabilidade com outros sistemas. Com a possibilidade de reutilização dos metadados oferecidos pelo Repositório, em especial com o protocolo OAI-PMH, outras bases podem coletar e alimentar seus catálogos. Como os registros do RI UNESP passam por diversos controles de qualidade no decorrer dos processos de inserção de dados, isso garante maior confiabilidade e qualidade dos dados disponibilizados.

Um exemplo disso pode ser dado pelo Repositório da Produção Científica do CRUESP (Figura 6), do qual o RI UNESP e os repositórios da USP e da UNICAMP fazem parte. Entretanto, para que os dados da USP, UNESP e UNICAMP fossem armazenados em um único ambiente, foi necessária a padronização de metadados, de modo a garantir um mínimo de interoperabilidade entre esses sistemas e apresentar resultados com maior qualidade em um único portal.



Figura 6 – Repositório da Produção Científica CRUESP

Fonte: Conselho de Reitores das Universidades Estaduais Paulistas (2017).

Na Figura 6, observa-se que pode ser realizada uma busca integrada, ou separadamente pela instituição. O sistema utilizado no caso do Repositório do CRUESP é o sistema de descoberta PRIMO. “Primo foi concebido para ser altamente personalizável para cada biblioteca, permitindo que a biblioteca para definir a sua própria indexação e apresentação regras, ponderações de relevância, bem como muitos outros detalhes de configuração” (BREEDING, 2011, p. 23). Ao realizar uma busca no Repositório CRUESP, é possível ver os metadados coletados e o *link* de acesso para o RI de origem, conforme apresentado no Quadro 4.

O mesmo processo ocorre com as teses e as dissertações defendidas na Unesp, que são coletadas pela Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), que é administrada pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), desde 2002. O IBICT coleta e armazena os metadados do RI UNESP das teses e dissertações e no metadado URL disponibiliza o *link* de acesso para a fonte original. O Quadro 5 apresenta um registro disponibilizado na BDTD a partir da coleta do RI UNESP.

Título:	A interação do usuário com catálogos bibliográficos on-line: investigação a partir da Teoria Fundamentada
Autor:	Bastos, Flávia Maria
Editor:	Universidade Estadual Paulista (UNESP)
Data de publicação:	2013
Descrição:	Pós-graduação em Ciência da Informação – FFC This research aims to identify [...] Esta pesquisa tem como objetivo identificar [...]
Assuntos:	Catálogos de Bibliotecas Online; Interação Humano-computador; Dados Bibliográficos Legíveis por Computador
Links	Repositório UNESP

Quadro 4 – Metadados disponibilizados pelo Repositório CRUESP

Fonte: Conselho de Reitores das Universidades Estaduais Paulistas (2017).

id	UNSP_73bf1bc898b12e6198bc4ddfeeee3c40
oai_identifier	oai:agregador.ibict.br:RI_UNESP:oai:repositorio.unesp.br:11449/103366
network_acronym	UNSP
network_name	UNESP
thumbnail	UNESP_cover.png
recordtype	Driver
spelling	A interação do usuário com catálogos bibliográficos on-line: investigação a partir da Teoria Fundamentada Bastos, Flávia Maria [UNESP] Vidotti, Silvana Aparecida Borsetti Gregorio [UNESP] Catálogos de bibliotecas online Interação humano-computador Dados bibliográficos legíveis por computador This research aims to identify [...] Esta pesquisa tem como objetivo identificar [...] 2014-06-11T19:32:41Z 2014-06-11T19:32:41Z 2013-10-03 info:eu-repo/semantics/doctoralThesis http://hdl.handle.net/11449/103366 000735122 000735122.pdf 33004110043P4 0000-0002-4216-0374 por info:eu-repo/semantics/openAccess 255 f.: il. Universidade Estadual Paulista (UNESP) Aleph BASTOS, Flávia Maria. A interação do usuário com catálogos bibliográficos on-line: investigação a partir da Teoria Fundamentada. 2013. 255 f. Tese (doutorado) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências de Marília, 2013. reponame:RI info:eu-repo/semantics/publishedVersion instname:UNESP
reponame	RI
collection	RI
instname	UNESP
institution	UNESP

language	Por
format	doctoralThesis
status	publishedVersion
eu_rights	openAccess
author	Bastos, Flávia Maria [UNESP]
spellingShingle	Bastos, Flávia Maria [UNESP] A interação do usuário com catálogos bibliográficos on-line: investigação a partir da Teoria Fundamentada Catálogos de bibliotecas online Interação humano-computador Dados bibliográficos legíveis por computador
author-letter	Bastos, Flávia Maria [UNESP]
author2	Vidotti, Silvana Aparecida Borsetti Gregorio [UNESP]
author2Str	Vidotti, Silvana Aparecida Borsetti Gregorio [UNESP]
title	A interação do usuário com catálogos bibliográficos on-line: investigação a partir da Teoria Fundamentada
title_short	A interação do usuário com catálogos bibliográficos on-line: investigação a partir da Teoria Fundamentada
title_full	A interação do usuário com catálogos bibliográficos on-line: investigação a partir da Teoria Fundamentada
title_fullStr	A interação do usuário com catálogos bibliográficos on-line: investigação a partir da Teoria Fundamentada
title_full_unstemmed	A interação do usuário com catálogos bibliográficos on-line: investigação a partir da Teoria Fundamentada
title_sort	A interação do usuário com catálogos bibliográficos on-line: investigação a partir da Teoria Fundamentada
publisher	Universidade Estadual Paulista (UNESP)
topic	Catálogos de bibliotecas online Interação humano-computador Dados bibliográficos legíveis por computador
topic_facet	Catálogos de bibliotecas online Interação humano-computador Dados bibliográficos legíveis por computador
description	This research aims to identify [...]
publishDate	2013
url	<a href="http://hdl.handle.net/11449/103366">http://hdl.handle.net/11449/103366</a>

identifier	000735122 000735122.pdf 33004110043P4 0000-0002-4216-0374
version	1566660467682705408
score	13.643.562

Quadro 5 – Registro disponibilizado pela BDTD

Fonte: Instituto Brasileiro de Informação em Ciência e Tecnologia (2017a).

Além da BDTD, o IBICT lançou o Portal Brasileiro de Publicações Científicas em Acesso Aberto oasisbr, que reúne artigos científicos, teses, dissertações e trabalhos em eventos, advindos dos repositórios brasileiros e portugueses. A iniciativa é análoga à da BDTD e amplia as possibilidades de localização de objetos digitais no âmbito nacional e português. O Quadro 6 apresenta um registro coletado do RI UNESP no portal OASIS.

id	UNSP_9ac2583cb3652a6b5ca604ab05262f7c
oai_identifier	oai:agregador.ibict.br:RI_UNESP:oai:repositorio.unesp.br:11449/144718
network_acronym	UNSP
network_name	UNESP
thumbnail	UNSP_cover.png
recordtype	Driver
spelling	Coleta automática para povoamento de repositórios digitais: conversão de registros utilizando XSLT Vidotti, Silvana Aparecida Borsetti Gregorio [UNESP] Bastos, Flavia Maria [UNESP] Grisoto, Ana Paula [UNESP] Arakaki, Felipe Augusto [UNESP] Ferreira, Juliano Benedito [UNESP] Reuse of metadata UNESP Institutional Repository XSLT stylesheets Automatic conversion Reuso de metadados Repositório Institucional UNESP Folhas de estilo XSLT Conversão automática The creation of digital repositories meets the proposed disclosure and increased visibility [...] A criação de repositórios digitais vem ao encontro das propostas de divulgação e aumento da visibilidade [...] 2016-11-28T13:37:26Z 2016-11-28T13:37:26Z 2016-11 info:eu-repo/semantics/conferenceObject <a href="http://www.ufpb.br/evento/lti/ocs/index.php/enancib2016/enancib2016/paper/view/4092">http://www.ufpb.br/evento/lti/ocs/index.php/enancib2016/enancib2016/paper/view/4092</a> 1676-546X <a href="http://hdl.handle.net/11449/144718">http://hdl.handle.net/11449/144718</a> 2016-01-21-vidotti.pdf por Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) info:eu-repo/semantics/openAccess 1-21 Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB) In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., 2016, Bahia, 2016. Anais... Bahia: ANCIB; UFBA, 2016, p. 1-21. reponame:RI info:eu-repo/semantics/publishedVersion instname:UNESP

reponame	RI
collection	RI
instname	UNESP
institution	UNESP
language	Por
format	conferenceObject
status	publishedVersion
eu_rights	openAccess
author	Vidotti, Silvana Aparecida Borsetti Gregorio [UNESP]
spellingShingle	Vidotti, Silvana Aparecida Borsetti Gregorio [UNESP] Coleta automática para povoamento de repositórios digitais: conversão de registros utilizando XSLT Reuse of metadata UNESP Institutional Repository XSLT stylesheets Automatic conversion Reuso de metadados Repositório Institucional UNESP Folhas de estilo XSLT Conversão automática
author-letter	Vidotti, Silvana Aparecida Borsetti Gregorio [UNESP]
author2	Bastos, Flavia Maria [UNESP] Grisoto, Ana Paula [UNESP] Arakaki, Felipe Augusto [UNESP] Ferreira, Juliano Benedito [UNESP]
author2Str	Bastos, Flavia Maria [UNESP] Grisoto, Ana Paula [UNESP] Arakaki, Felipe Augusto [UNESP] Ferreira, Juliano Benedito [UNESP]
title	Coleta automática para povoamento de repositórios digitais: conversão de registros utilizando XSLT
title_short	Coleta automática para povoamento de repositórios digitais: conversão de registros utilizando XSLT
title_full	Coleta automática para povoamento de repositórios digitais: conversão de registros utilizando XSLT
title_fullStr	Coleta automática para povoamento de repositórios digitais: conversão de registros utilizando XSLT

title_full_unstemmed	Coleta automática para povoamento de repositórios digitais: conversão de registros utilizando XSLT
title_sort	Coleta automática para povoamento de repositórios digitais: conversão de registros utilizando XSLT
publisher	Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB)
topic	Reuse of metadata UNESP Institutional Repository XSLT stylesheets Automatic conversion Reuso de metadados Repositório Institucional UNESP Folhas de estilo XSLT Conversão automática
topic_facet	Reuse of metadata UNESP Institutional Repository XSLT stylesheets Automatic conversion Reuso de metadados Repositório Institucional UNESP Folhas de estilo XSLT Conversão automática
description	The creation of digital repositories meets the proposed [...]
publishDate	2016
url	<a href="http://www.ufpb.br/evento/lti/ocs/index.php/enancib2016/enancib2016/paper/view/4092">http://www.ufpb.br/evento/lti/ocs/index.php/enancib2016/enancib2016/paper/view/4092</a> <a href="http://hdl.handle.net/11449/144718">http://hdl.handle.net/11449/144718</a>
identifier	1676-546X 2016-01-21-vidotti.pdf
_version_	1566681439013961728
score	150.862.055

Quadro 6 – Registro do OASIS

Fonte: Instituto Brasileiro de Informação em Ciência e Tecnologia (2017b).

Outro caso de integração a partir dos dados do RI UNESP, é um acordo de cooperação entre a FAPESP e a UNESP. O acordo permitirá o envio dos dados das produções científicas que tiveram apoio financeiro da FAPESP, a partir dos metadados de agência de fomento e o número de processo. O acordo beneficiará a FAPESP, que terá condições de acompanhar a evolução e retorno dos trabalhos por ela financiados, verificando, por exemplo, se os resultados de um projeto

financiado foram publicados. A Unesp terá um controle maior dos projetos que possuem financiamentos e poderá desenvolver, por exemplo, indicadores para melhorar a gestão. A população em geral ganhará com a transparência da divulgação dos resultados, podendo acompanhar o projeto desde o seu início até os resultados finais. A Figura 7 apresenta um projeto financiado pela FAPESP, constando a informação de uma publicação, no caso uma dissertação.

Figura 7 – Projeto financiado pela FAPESP

Fonte: Fundação de Amparo à Pesquisa do Estado de São Paulo (2017).

Em julho de 2015, a Unesp anunciou a adesão ao *Open Research and Contributors Identification* (ORCID). O ORCID é um identificador único para pesquisadores e contempla 16 caracteres que possuem abrangência mundial. Ele permite a desambiguação de nomes, garantindo que a publicação do pesquisador seja inequívoca.

Com um identificador internacional, a Unesp tem maior controle na identificação da produção científica de seus pesquisadores. Assim, quando o pesquisador publicar um artigo em um periódico e informar o seu ORCID ID, esse número persistirá nos metadados do registro da revista ou em outras bases de dados como Scopus, SciELO e WoS. Quando atualizar sua base de dados, coletando os registros dessas bases indexadoras, a partir do ORCID ID, a Unesp saberá se o autor é um pesquisador da Unesp ou de outra instituição, o que facilita as coletas automáticas. O pesquisador terá sua produção coletada automaticamente para o RI UNESP, reduzindo a ambiguação do seu nome com o de outros pesquisadores.



O pesquisador precisa autorizar a Unesp a alimentar seu ORCID ID com a sua produção científica existente no Repositório. Uma vez autorizado, o Repositório mapeia a produção do docente no Repositório e envia esses dados para o ORCID ID do pesquisador, conforme apresentado na Figura 8.



Figura 8 – Fluxo de integração RI UNESP com ORCID

Fonte: Autoria própria (2016).

Com o ORCID ID, o RI UNESP terá maior controle dos registros de autoridade de pessoas em um menor espaço de tempo, verificando se as informações dos pesquisadores estão corretas e padronizadas, como nomes abreviados, troca de sobrenome depois de matrimônio, entre outras inconsistências. Após a importação dos registros, alguns metadados ficam disponíveis na base do ORCID com um *link* direcionado para RI UNESP e com a informação de que o registro foi coletado da Universidade Estadual Paulista Júlio Mesquita Filho, conforme a Figura 9.

Com o crescimento do repositório, as demandas decorrentes de dúvidas e as perguntas via e-mail aumentaram proporcionalmente. O RI UNESP recebia diversos e-mails com várias demandas, e muitas vezes os e-mails eram confusos, dificultando uma resposta rápida que atendesse corretamente ao usuário. Para solucionar esse problema, foi adquirido um ambiente para auxiliar no gerenciamento de questões dos usuários da *Online Computer Library Center* (OCLC), o *Question Point*.

O *Question Point* possui duas possibilidades de atendimento ao usuário, o chat e o formulário de pergunta. No *chat*, foram concentradas questões mais pontuais como problemas em registros, dúvidas sobre o autoarquivamento e ORCID.

Com a nova dinâmica, as respostas são dadas em tempo real, agilizando o atendimento aos usuários. Para outras questões ou quando o *chat* está indisponível, o usuário encaminha uma mensagem pelo formulário de pergunta que é gerenciado pelo sistema e, assim que possível, as perguntas são respondidas. O histórico das respostas fica armazenado por algum tempo no sistema e possibilita que outros bibliotecários acompanhem o que está sendo perguntado e respondido.

Coleta automática para povoamento de repositórios digitais:  
conversão de registros utilizando XSLT

2016-11-01 | conference-paper

HANDLE: 11449/144718

URL: <http://hdl.handle.net/11449/144718>

---

#### Language

Portuguese

#### URL

<http://hdl.handle.net/11449/144718>

#### Citation (formatted-unspecified)

In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., 2016, Bahia, 2016. Anais... Bahia: ANCIB; UFBA, 2016, p. 1-21.

#### Description

The creation of digital repositories meets the proposed disclosure and increased visibility of the institutions and universities that aims to broaden the dissemination of developed research activities and production in the community. The UNESP Institutional Repository began in 2013, is part of the global context of universities that

#### Contributor

Vidotti, Silvana Aparecida Borsetti Gregorio (author, first)

Bastos, Flavia Maria (author, additional)

Grisoto, Ana Paula (author, additional)

Arakaki, Felipe Augusto (author, additional)

Ferreira, Juliano Benedito (author, additional)

#### Created

2017-02-16

Figura 9 – Registro coletado do RI UNESP para ORCID

Fonte: Vidotti (2017).

## CONSIDERAÇÕES FINAIS

Desde a implantação do RI UNESP, em outubro de 2013, foram observados diversos benefícios para a Unesp. Considera-se que o repositório proporcionou o aumento da visibilidade da produção científica da Unesp, pois, de junho de 2014 até junho de 2017, o RI UNESP recebeu mais de 5 milhões de visitantes e teve mais de 9 milhões de páginas visualizadas, sendo que, em 2017, recebeu mais de 20 mil visualizações por dia. Em pouco tempo desde sua criação, o RI UNESP obteve resul-

tados significativos no principal *ranking* que avalia os repositórios, o *Web Ranking of World Repositories*, saindo da 42ª posição nacional em julho de 2014, para ocupar a 5ª posição em janeiro de 2017.

As atividades direcionadas à padronização dos metadados, bem como o controle do nome da Universidade e dos autores também possibilitaram maior visibilidade da produção da Unesp nos *rankings*, uma vez que o aprimoramento no tratamento da informação da produção da Unesp passou a ampliar a possibilidade de coleta pelos rankings nacionais e internacionais, o que contribuiu para melhor posicionamento da Universidade.

O trabalho desenvolvido pela equipe do RI UNESP tem se tornado referência nacional, tanto no âmbito da atuação profissional, quanto para o desenvolvimento de pesquisas e inovação. No campo profissional, diversas instituições de ensino superior solicitaram transferência da metodologia e dos procedimentos adotados pelo RI UNESP. Dentre elas, é de citar a Universidade Estadual de Campinas (UNICAMP), a Universidade Federal de São Paulo (UNIFESP), a Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), o Instituto de Pesquisas Energéticas e Nucleares (IPEN), o Instituto de Engenharia Nuclear (IEN), a Universidade Federal do Rio de Janeiro (UFRJ) e a Universidade Tecnológica Federal do Paraná (UTFPR).

No âmbito acadêmico, o trabalho que aborda a metodologia do RI UNESP foi destaque no Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), principal evento de Pós-Graduação da área de Ciência da Informação no Brasil, no ano de 2016, ratificando a importância do trabalho desenvolvido pela Unesp.

Com isso, novos desafios estão sendo projetados para o RI UNESP, como a implantação de um sistema para gerenciar a publicação de dados de pesquisa, o desenvolvimento de indicadores da produção científica por departamentos, áreas, unidades, instituições e, ainda, a construção de redes de colaboração entre os autores e as instituições.

A representação temática dos recursos bibliográficos é outro item de inovação pela perspectiva de recuperação com precisão dos documentos. Nesse sentido, o estudo de uma política de indexação com o vocabulário dos registros torna-se estratégico à medida que poderiam ser incorporados vocabulários controlados de uso internacional.

## REFERÊNCIAS

ASSUMPTÃO, F. S. et al. A conversão de registros na implantação de repositórios institucionais: o caso do Repositório Institucional UNESP. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 18., 2014, Belo Horizonte. **Anais...** Belo Horizonte: UFMG, 2014. p. 1-16. Disponível em: <<https://repositorio.unesp.br/handle/11449/123645>>. Acesso em: 15 maio 2017.

BAPTISTA, A. A.; MACHADO, A. B. Um gato preto num quarto escuro: falando sobre metadados. **Revista de Biblioteconomia de Brasília**, v. 25, n. 1, p. 77-90, 2001. Disponível em: <[www.brapci.ufpr.br/download.php?dd0=12263](http://www.brapci.ufpr.br/download.php?dd0=12263)>. Acesso em: 26 jun. 2012.

BREEDING, M. Beyond the ILS : a new generation of library platforms. In: IGLESIAS, E. (Org.). **Robots in academic libraries: advancements in library automation**. 2011. Cap. 2. p. 13-32. Disponível em: <<http://www.igi-global.com/chapter/beyond-ils-new-generation-library/76457>>. Acesso em: 22 fev. 2013.

CONSELHO DE REITORES DAS UNIVERSIDADES ESTADUAIS PAULISTAS. **Repositório da Produção Científica CRUESP**. Disponível em: <[www.repositorio.cruesp.sp.gov.br](http://www.repositorio.cruesp.sp.gov.br)>. Acesso em: 23 maio 2017.

COYLE, K.; BAKER, T. **Guidelines for Dublin Core application profiles**. DCMI, 2009. Disponível em: <<http://dublincore.org/documents/profile-guidelines/>>. Acesso em: 16 nov. 2014.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. **Repositório digital como ambiente de inclusão digital e social para usuários idosos**. 2017. Disponível em: <<http://www.bv.fapesp.br/pt/bolsas/105004/repositorio-digital-como-ambiente-de-inclusao-digital-e-social-para-usuarios-idosos/>>. Acesso em: 23 maio 2017.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **A interação do usuário com catálogos bibliográficos on-line: investigação a partir da Teoria Fundamentada**. Disponível em: <[http://bdtb.ibict.br/vufind/Record/UNSP\\_73bf1bc898b12e6198bc4ddfeeee3c40](http://bdtb.ibict.br/vufind/Record/UNSP_73bf1bc898b12e6198bc4ddfeeee3c40)>. Acesso em: 23 maio 2017a.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Coleta automática para povoamento de repositórios digitais: conversão de registros utilizando XSLT**. Disponível em: <[http://oasisbr.ibict.br/vufind/Record/UNSP\\_9ac2583cb3652a6b5ca604ab05262f7c](http://oasisbr.ibict.br/vufind/Record/UNSP_9ac2583cb3652a6b5ca604ab05262f7c)>. Acesso em: 23 maio 2017b.

NILSSON, M.; BAKER, T.; JOHNSTON, P. **The Singapore Framework for Dublin Core application profiles**. DCMI, 2008. Disponível em: <<http://dublincore.org/documents/singapore-framework/>>. Acesso em: 18 jan. 2014.

REESE, T. **MarcEdit (5.3)** [software]. [S.L.]: MarcEdit Development, 2016.

REPOSITÓRIO INSTITUCIONAL UNESP. **DSpace OAI-PMH data provider**. 2017. Disponível em: <<https://repositorio.unesp.br/oai/request?verb=Identify>>. Acesso em: 23 maio 2017.

UNIVERSIDADE ESTADUAL PAULISTA. Portaria UNESP n. 88, de 28 de fevereiro de 2013. **Diário Oficial do Estado de São Paulo**, São Paulo, SP, 01 mar. 2013. Caderno Executivo 1, p. 47. Disponível em: <[https://www.imprensaoficial.com.br/DO/BuscaDO2001Documento\\_11\\_4.aspx?link=/2013/executivo%2520secao%2520i/marco/01/pag\\_0047\\_FID5L3M88QPU4eANIAOIQF3FK85.pdf&pagina=47&data=01/03/2013&caderno=Executivo%20I&paginaordenacao=100047](https://www.imprensaoficial.com.br/DO/BuscaDO2001Documento_11_4.aspx?link=/2013/executivo%2520secao%2520i/marco/01/pag_0047_FID5L3M88QPU4eANIAOIQF3FK85.pdf&pagina=47&data=01/03/2013&caderno=Executivo%20I&paginaordenacao=100047)>. Acesso em: 23 maio 2017.

VIDOTTI, S. A. B. G. **Currículo ORCID**. 2017. Disponível em: <<http://orcid.org/0000-0002-4216-0374>>. Acesso em: 23 maio 2017.

VIDOTTI, S. A. B. G. et al. Coleta de dados a partir dos currículos da plataforma Lattes: procedimentos utilizados no Repositório Institucional UNESP. **Ponto de Acesso**, Salvador, v. 9, n. 3, p. 117-132, dez. 2015a. Disponível em: <<https://portalseer.ufba.br/index.php/revistaici/article/view/15164>>. Acesso em: 15 maio 2017.

VIDOTTI, S. A. B. G. et al. Reutilização de metadados para o povoamento de um repositório institucional: procedimentos aplicados no Repositório Institucional UNESP. In: INTERNATIONAL CONFERENCE ON *Dublin Core* & METADATA APPLICATIONS (DC-2015), 15., 2015, São Paulo. **Proceedings...**, 2015b. Disponível em: <<https://repositorio.unesp.br/handle/11449/127972>>. Acesso em: 24 maio 2017.

VIDOTTI, S. A. B. G. et al. Coleta automática para povoamento de repositórios digitais: conversão de registros utilizando XSLT. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 9, n. 2, set./dez. 2016. Disponível em: <<http://inseer.ibict.br/ancib/index.php/tpbci/article/view/250>>. Acesso em: 15 maio 2017.



# O PAPEL DOS REPOSITÓRIOS DIGITAIS NA CURADORIA DE DADOS DE PESQUISA

Luis Fernando Sayão









## INTRODUÇÃO

Uma revolução de contornos bem nítidos está ocorrendo em todos os domínios dos empreendimentos humanos. Identificada como um fenômeno do nosso tempo, ela é caracterizada por um vertiginoso aumento na quantidade de informação digital disponível; pelo dinamismo tanto dos propósitos quanto da oferta de tecnologias e ferramentas voltadas para a utilização dessas informações; e, ainda, pela multiplicidade de agentes capazes de manipular esses estoques informacionais digitais. As condicionantes que engendram esse cenário disruptivo são muitas, todavia, as tecnologias computacionais e de rede que permitem a criação, captura, transmissão, cópia, compartilhamento e armazenamento de uma quantidade massiva de informação de maneira fácil e a um custo muito baixo, estão no centro dessa revolução da informação digital, que se estende de forma contundente ao mundo da ciência (NATIONAL RESEARCH COUNCIL, 2015).

A partir do final do século passado, dados digitais coletados ou gerados por novos instrumentos, satélite, sensores remotos e simulações computadorizadas deixam muito rapidamente de ser meros subprodutos das atividades de pesquisa e se tornam um foco de grande interesse para todo o mundo científico, delineando novos padrões metodológicos para a pesquisa científica e permitindo tipos de pesquisa até então não possíveis de realizar. A grande disponibilidade de dados – o *big data* científico – transformou muitas disciplinas em áreas baseadas intensivamente em dados, como biologia molecular, biodiversidade, estudos do ecossistema e geografia; áreas híbridas como bioinformática, geoinformática e ciências geoespaciais surgem em resposta à imensa quantidade de dados e caracterizam um novo tipo de ciência onde a formação de bases de dados se torna um fim em si mesmo (NATIONAL RESEARCH COUNCIL, 2015; BOULTON et al., 2012).

Em ambientes virtuais, pesquisas interdisciplinares, colaborativas e geograficamente distribuídas, geram e usam vasta quantidade de dados, recriam os percursos metodológicos de identificação de novos fenômenos e de formulação de hipóteses; redesenham os processos de avaliação, validação e reprodutibilidade das pesquisas; e inauguram novas formas de socialização entre pesquisadores, refazendo o ciclo tradicional de comunicação científica. Esses fatos definem padrões inéditos de investigação científica coletivamente conhecida como e-Science. As pesquisas nesses ambientes se desenrolam apoiadas em arcabouços altamente tecnológicos chamados de ciberinfraestruturas ou e-infraestruturas, que formam a espinha dorsal de apoio aos serviços da e-Science (MARTINEZ-URIBE;

MACDONALD, 2008). Este cenário de grandes novidades para o mundo da ciência, identificado como **quarto paradigma científico**, estabelece novos métodos de investigação científica, onde a análise de coleções de dados abertos e curados por especialistas complementa e estende os métodos tradicionais como simulação, observação e experimentação teórica (GRAY et al., 2002).

O crescente interesse pelos dados digitais coletados ou gerados pelas atividades de pesquisa na última década – por toda a sociedade e não só pelos segmentos acadêmicos e científicos – criou uma demanda por estruturas organizacionais, tecnológicas e por capital humano que pudessem dar conta da gestão, sustentabilidade e análise desses novos ativos informacionais. Isso porque os dados são objetos digitais complexos que “não são compostos somente por *bytes*, mas também por definições de fluxos de trabalho, parâmetros computacionais, configurações de ambientes e muito mais” (VALLE, 2017); além disso, a tecnologia digital cria e destrói seus artefatos: a mesma tecnologia que permite a criação, manipulação e torna acessível, de muitas maneiras, quantidades imensas de informação coloca em perigo a longevidade dos objetos que ela cria e exige uma gestão muito diferente da praticada no mundo baseado em papel (LEE; TIBBO, 2007).

A natureza complexa e heterogênea dos dados de pesquisa implica numa gestão dinâmica que vai além do armazenamento seguro e da disponibilização na web. Essa gestão, cujo conjunto de procedimentos é chamado de curadoria de dados, adiciona valor aos dados, aumentando, dessa forma, o seu potencial de utilização atual e futuro e a probabilidade de reuso e citação por outros pesquisadores provenientes de domínios disciplinares diferentes da área para a qual os dados foram originalmente coletados ou gerados.

Para apoiar a execução dos processos de gestão e de curadoria de dados, é necessário um arcabouço tecnológico e organizacional que compreenda todo o ciclo de vida dos dados. No centro desse arcabouço estão os repositórios digitais de dados de pesquisa que, por muitas razões e demandas, rapidamente se tornam parte essencial da infraestrutura mundial de pesquisa.

O objetivo do presente ensaio é discutir rapidamente o papel dos repositórios de dados enquanto elemento essencial do fluxo de gestão de dados de pesquisa, e como parte da infraestrutura necessária à implantação de serviços de curadoria de dados integrados aos sistemas de informação acadêmicos, principalmente às bibliotecas de pesquisa. Toma-se como metodologia a análise de algum dos principais autores da área, e as discussões em torno dos elementos condicionantes dos repositórios: dados de pesquisa, objetos informacionais, reuso, gestão, preservação e curadoria de dados de pesquisa.

## NATUREZA FRÁGIL E HETEROGÊNEA DOS DADOS

Os dados de pesquisa estão se tornando a moeda proverbial da ciência, porém, contrariamente à maioria dos sistemas monetários, eles podem fluir em imensa abundância. Isso acontece principalmente porque os dados de pesquisa provêm de uma infinidade de fontes: animais correndo nas savanas africanas ou aprisionados em zoológicos; fenômenos da natureza como marés e vulcões; instrumentos, sensores, modelos matemáticos e simulações da natureza; livros, pensamentos e ideias. É exatamente a abundância dos dados de pesquisa que desafia os profissionais de informação, de computação e os cientistas de dados a definir estratégias que possam controlar o fluxo de dados provenientes das atividades acadêmicas e de pesquisa, e preservar as evidências únicas para uso futuro. Isso significa assegurar a resiliência dos dados digitais, equacionando os problemas de como o significado, integridade, autenticidade e proveniência dos dados digitais gerados pelos atuais **pesquisadores serão capturados e transmitido no tempo e no espaço para o reuso de outros pesquisadores** (JOHNSTON, 2017).

A natureza diversificada dos dados de pesquisa torna um enigma – mesmo para o pesquisador, mesmo para os especialistas que lidam no dia a dia com a gestão de dados – definir com precisão o que é dado de pesquisa de uma forma transversal aos infinitos domínios disciplinares da fragmentada ciência contemporânea. O termo **dado de pesquisa** tem uma amplitude de significados, que vão se transformando de acordo com domínios científicos específicos, objetos de pesquisas, metodologias de geração e coleta de dados e muitas outras variáveis.

Na sua análise sobre a natureza dos dados, Borgman (2007, p. 119) ressalta que a “Informação é um conceito complexo com centenas de definições [...]. Dado [por sua vez] é um conceito simples com poucas definições, porém sujeito a muitas e diferentes interpretações”. O que dificulta, portanto, atribuir uma definição consensual a dado de pesquisa é o fato idiossincrático que ele pode ser muitas coisas diferentes para pessoas e circunstâncias diferentes. Isso acontece porque dado de pesquisa é dependente de interpretação e de contextos específicos que o identificam como um produto de pesquisa. O status dos dados como fato ou evidência de uma ocorrência de valor científico é determinado pela pessoa que coleta, gerencia e usa esses dados, completa Borgman (2007). Esse mesmo relativismo é ratificado por Nielsen e Hjørland (2014, p. 225): “Dados são sempre registrados tomando como base algum interesse, perspectiva, tecnologia e prática que determinam seus significados e utilidades em diferentes contextos”.

Continuando sua análise, Borgman (2007) alinha uma série de exemplos que ilustram as afirmações acima:

- a) uma sequência de bits proveniente de um sensor sísmico é dado de pesquisa para os sismólogos;
- b) amostras de rochas são dados de pesquisa para um geomorfologista;
- c) conversas gravadas são dados de pesquisa para sociólogos;
- d) inscrições em cuneiformes são dados de pesquisa para quem estuda linguagens do Oriente Próximo.

Porém, os cuneiformes podem ser também dados para o arqueólogo e para o ambientalista que buscam padrões climáticos históricos; de forma análoga, os dados sísmicos podem ser úteis para biólogos que estudam comportamento animal.

Nestes últimos exemplos, os dados originalmente considerados em um domínio disciplinar são submetidos a novos contextos e ressignificados por novos olhares. Essa possibilidade torna os dados de pesquisa um recurso informacional de grande valor e importância; em contrapartida, exige uma gestão mais criteriosa e sofisticada, capaz de revelar novas dimensões de informação e conhecimento e transportá-las para o futuro.

O dado de pesquisa é caracterizado pela heterogeneidade em termos de origens, formatos, tecnologias, áreas disciplinares, para citar algumas possíveis variáveis. Primordialmente associados com determinadas áreas científicas, onde desde sempre foram considerados recursos informacionais de primeira grandeza, dados digitais de pesquisa podem ser identificados em qualquer disciplina e em muitas formas (números, figuras, vídeos, textos, softwares, entre outros); com diferentes níveis de agregação e de processamento, como dados crus ou primários (como números coletados diretamente de um instrumento, dados intermediários e dados processados); podem ser a agregação de múltiplas fontes ou o produto de modelos, simulações ou processos de visualização. Dados das humanidades digitais podem incluir textos digitalizados ou nascidos digitais, bibliotecas de imagens digitais e modelos 3D, tais como os usados para reconstrução histórica de sítios antigos. Os cientistas sociais também produzem um volume considerável de dados, que incluem dados de levantamentos, pesquisa de opinião pública, mapas com informações georreferenciadas, observações e indicadores, tais como os capturados *on-line* por sensores ou vídeos de atividades e interações humanas complexas (JOHNSTON, 2017; MARTINEZ-URIBE; MACDONALD, 2008).

O relatório da National Science Board (2005) categoriza os dados de pesquisa em:

- a) **dados experimentais** - resultados de estudos em ambientes controlados de laboratórios;

- b) **dados computacionais** - produtos da execução de modelos computacionais que simulam uma dada realidade;
- c) **dados observacionais** - resultados de observações de fenômenos que se desenrolam em lugares e tempos específicos.

Essa categorização ajuda a definir o grau de gestão e de curadoria necessários para manter a usabilidade de cada tipo de dados. Por exemplo, os dados observacionais precisam ser preservados *ad aeternum*, pois não há como coletá-los novamente e, portanto, requerem uma gestão de longo prazo mais intensa e cuidadosa; por outro lado, no âmbito da simulação, os modelos computacionais é que precisam ser preservados e não os dados resultantes da execução de códigos gerados a partir desses modelos (SAYÃO; SALES, 2015).

O governo é outra fonte de grandes volumes de informações digitais. Dados provenientes de recenseamento, registros médicos, seguro social, registros da vida pública e privada geram dados que são críticos para as pesquisas nas áreas de saúde, ciências, ciências sociais e humanidades (NATIONAL RESEARCH COUNCIL, 2015; BORGMAN, 2007). Porém, o compartilhamento de coleções de dados contendo informações pessoais coloca desafios importantes para a governança da informação e para a proteção da confidencialidade, visto que os processos de *anonimização* não garantem a segurança dos registros pessoais (BOULTON et al., 2012).

Ao largo do mundo da pesquisa e dos setores governamentais, o setor de negócios, a indústria e o comércio produzem o *big data* que pode ser usado, entre muitas outras coisas, para endereçar questões científicas que melhorem a compreensão do comportamento humano, porém trazem como resultado a emergência de uma crescente – e algumas vezes nefasta – economia baseada no comércio transnacional de dados de negócios (JOHNSTON, 2017).

Apesar da sua importância, os dados são extremamente frágeis, especialmente em relação à passagem do tempo. Eles dependem de uma hierarquia de tecnologias em constante – e rápida – transformação: hardware, mídias de armazenamento, sistemas operacionais, software de aplicação e *middleware*; dependem também de conhecimento tácito externo aos dados. Sem os devidos cuidados, as informações codificadas como uma insubstancial cadeia de *bits* se tornarão rapidamente inacessíveis, sem confiança e sem significado (LORD; MACDONALD, 2004).

## DADO DE PESQUISA COMO OBJETO INFORMACIONAL

Quase tudo que é gerado e coletado no ambiente de pesquisa pode ser considerado dado de pesquisa. Esse fato coloca a importância de examinar o conceito de dados de pesquisa ancorado por um sistema de informação que pode formalizá-lo como objeto de informação por meios de fluxos de trabalho apropriados e processos de representação, contextualização e ressignificação que os tornem passíveis de servir a diferentes disciplinas ou comunidades científicas.

A criação de banco de dados de coleções de dados de pesquisa e a publicação dessas coleções na web, não são medidas suficientes para dotar esses ativos de usabilidade e capacidade de transmitir conhecimento. O significado nem sempre está associado diretamente aos dados, pois eles não são tão autodescritos como um artigo acadêmico ou um livro, cujos significados estão contidos nos próprios artefatos. Sem uma documentação associada, os dados são apenas cadeias de números, variáveis, imagens, fragmentos de texto, áudio ou vídeo, ou outro elemento discreto qualquer. No caso de objetos digitais, eles são inúteis sem um software específico que decodifique sua cadeia de bits (BORGMAN, 2007). “Dados de pesquisa não têm valor sem seus metadados e documentação apropriada que descrevam seus contextos e as ferramentas usadas para criá-los, armazená-los, adaptá-los e analisá-los” (KINDLING; SCHIRMBACHER, 2013, p. 4). Para entender os dados, os usuários precisarão de metadados que explicitem como os instrumentos usados foram projetados e construídos; quando, onde e como os dados foram gerados ou coletados; e uma cuidadosa descrição dos estágios de processamento que geraram os produtos derivados dos dados, que são tipicamente usados para análise científica posterior (GRAY et al., 2002).

Os dados digitais de pesquisa são objetos digitais complexos que precisam de informação de representação semântica e estrutural que se manifestam por um conjunto amplo de metadados de vários tipos: descritivos, administrativos, técnicos, estruturais e disciplinares – que são os que descrevem as especificidades da área, como instrumentos, parâmetros, unidades e muito mais; e por uma documentação variada que inclui, por exemplo, caderno de pesquisa, livros de códigos, dicionários de dados, roteiro de entrevista, anotações, projetos e arquivos **leia-me**.

A utilidade dos dados depende também da capacidade dos pesquisadores em gerenciá-los, agregá-los e analisá-los. Nessa direção, algoritmos e ferramentas de mineração de dados, dispositivos de visualização, portais e interfaces para usuários desempenham um papel importante na otimização da usabilidade dos dados, e são elementos cruciais na aceleração das pesquisas. A área de estudos em visualização de dados, nesse contexto, é de grande importância, pois estuda a

transformação de dados, informação e conhecimento em representações visuais interativas, apoiando os pesquisadores na exploração, compreensão e análise dos dados (LIU, 2014). “Visualização de dados e ferramentas de mineração de dados são usadas para identificar e interpretar padrões que vão das Ciências Ambientais até a História da Arte”, completa Borgman (2007, p. 117).

Num ecossistema de dados, muitas inter-relações existem entre artigos, dados, software e outros recursos usados na pesquisa científica (MAYERNIK et al., 2016). O valor informacional dos dados se amplia pela contextualização que se realiza por meio do link com outros recursos. Dados e documentos acadêmicos têm mais valor quando interconectados do que quando isolados; eles podem ser conectados por links – que podem portar valor semântico – formando um arcabouço de informação acadêmica que torna possíveis novas formas de dados e de documentos mais criativos, interativos, distribuídos e colaborativos, que facilitam as pesquisas multidisciplinares e o aprendizado (BORGMAN, 2007; SALES, 2014). Além do mais, os artigos de periódicos vinculados por links aos dados em que eles estão baseados, são mais fundamentados, mais citados e possuem um nível de confiabilidade maior. O desenvolvimento das tecnologias de dados ligados (*linked data*) cria novas informações por meio de uma integração intensiva de diferentes coleções de dados, que potencialmente ampliam as possibilidades de enfoques automatizados para análise de dados (BOULTON et al., 2012).

Os elementos de representação e contextualização que são atribuídos pelos processos de gestão de dados resumem as ações necessárias para tornar os dados passíveis de serem descobertos, acessíveis, compreensíveis e compartilháveis ao longo do tempo. Porém, a gestão de dados será mais eficiente se for planejada no início do processo de pesquisa (LEIDEN UNIVERSITY, 2015); nesse sentido, o documento de planejamento de gestão de dados constitui uma formalização dessas ações e um compromisso entre os diversos atores envolvidos – pesquisadores, profissionais de informação, instituição e órgão de fomento à pesquisa (SAYÃO; SALES, 2015).

No ciclo de vida da gestão de dados, se incluem as ações mais contundentes e dinâmicas de adicionar valor aos dados para o reuso, chamada de curadoria de dados. Entretanto, para fazer parte da cadeia de geração de novos conhecimentos, como objetos informacionais, os dados precisam passar por processos de controle de qualidade, autenticação e validação que os habilite como objetos confiáveis para reuso de outros pesquisadores.

## REUSO: O OBJETIVO FINAL DA CURADORIA DE DADOS DE PESQUISA

Assim como geram e coletam novos dados digitais, os cientistas e acadêmicos começam a confiar nos conteúdos digitais criados por outros pesquisadores para darem prosseguimento às suas pesquisas (ABBOUT, 2008), caracterizando um padrão de compartilhamento de produtos de pesquisa conhecido como reuso de dados. Nessa direção, uma grande parcela das informações digitais está sendo usada ou reusada de uma forma não prevista quando a informação foi coletada ou gerada.

O reuso de dados coloca em destaque o fato de que as informações digitais se movem facilmente através dos limites temporais, como os diários de bordos de navios do século XVII digitalizados e depois analisados por climatologista do século XX; as informações digitais se movem também através dos limites setoriais quando os epidemiologistas examinam dados comerciais sobre consumo em busca de remédios para a gripe; a informação digital ignora também as fronteiras disciplinares quando pesquisadores em bioinformática combinam coleções de dados originados no domínio da biologia, genética e engenharia (NATIONAL RESEARCH COUNCIL, 2015). Isso significa que os dados precisam estar disponíveis e preparados para futuras pesquisas. “Mesmo hoje, alguns cientistas estão buscando registros astronômicos de 50 a 500 anos atrás – a maioria para realizar comparações temporais” ratifica Gray et al. (2002, p. 6).

Uma vez capturadas em formatos legíveis por computador, apropriadamente documentadas e armazenadas em plataformas conectadas à internet, as coleções de dados de pesquisa podem ser distribuídas em escala planetária. Os dados de pesquisa podem ser compartilhados e combinados de muitas formas, engendrando novos serviços e produtos e alimentando formas inéditas de cooperação entre pesquisadores na busca por novos conhecimentos.

“Todavia, gerar, capturar e gerenciar coleções de dados para uso corrente e futuro não é um desafio trivial. As atividades que são locais e tácitas se tornam globais e explícitas”, revela Borgman (2007, p. 118), o ponto mais crítico do ciclo de vida dos dados. Tornar um conteúdo criado para uma audiência útil para outra é um problema complexo, porque cada disciplina tem seu próprio vocabulário, estrutura de dados e práticas de pesquisa e formula questões de forma distinta usando sua própria terminologia. Isso coloca um desafio importante para os serviços de curadoria: criar descrições e representações, ferramentas e serviços que tornem viável o compartilhamento entre diferentes audiências (BORGMAN, 2007).



O dinamismo no uso e reuso da informação digital desperta uma alta demanda pelo estabelecimento de serviços de curadoria e expertises nessa área. “O desafio curatorial de interoperabilidade e acessibilidade são grandes quando os usos da informação digital são tão dispersos e fluidos” (NATIONAL RESEARCH COUNCIL, 2015, p. 9). A probabilidade de uma coleção de dados ser reusada no futuro por outras audiências, estabelece o critério mais simples de valor para a coleção; embora não seja algo simples, a partir daí pode-se estimar se vale pena arquivá-la por longo prazo (BORGMAN, 2007).

## **CURADORIA DE DADOS DE PESQUISA**

Para que o reuso de dados se constitua numa ponte confiável entre pesquisadores que estão dispersos geográfica e temporalmente, é necessário que esses conteúdos estejam submetidos a uma gestão ativa que assegure, além da capacidade de os dados serem interpretados, níveis adequados de integridade, autenticidade e de proveniência, que documente a origem e as subseqüentes transformações sofridas por eles. Os fluxos de trabalho necessários para isso fazem parte da curadoria de dados, que é identificada como um serviço intelectual e técnico que integra os processos mais amplos de gestão de dados de pesquisa. A curadoria de dados tem como objetivo imediato adicionar valor às coleções de dados tendo como perspectiva a promoção e reuso desses ativos informacionais por longo prazo.

Em termos gerais, o substantivo curadoria denota a seleção, o cuidado e a preservação de coleções de objetos de arte estáveis, como pintura e esculturas, livros e manuscritos raros e únicos; espécimes naturais e físicos importantes, e outros artefatos considerados de valor cultural ou científico. Em termos espaciais, a curadoria ocorre em contextos organizacionais relativamente limitados como bibliotecas, arquivos, museus, galerias de arte, herbários e instituições similares (NATIONAL RESEARCH COUNCIL, 2015). Além do mais, curadoria visa também promover a disponibilidade dos objetos para audiências apropriadas. Na transição para o mundo digital, a curadoria assume novos e importantes desafios, especialmente quando se consideram: as imensas e sempre crescentes quantidades de conteúdos digitais constituídos, em grande parte, de objetos complexos e heterogêneos, que dependem de tecnologias específicas e pouco duradouras, muitas vezes distribuídos globalmente e submetidos a um constante processo de inovação e atualização; a necessidade de uma gestão contínua que se desenrole num contexto de usos, culturas e tecnologias em ininterrupta mudança; a grande diversidade de contextos organizacionais em que a curadoria de conteúdos digitais ocorre; e uma audiência que pode ser indefinida e localizada no futuro.

Esse cenário coloca em evidência a necessidade da ampliação do conceito de curadoria para que compreenda infindável diversidade dos artefatos digitais e de seus usuários, e os ambientes em mutação onde eles precisam ser gerenciados.

Em termos de significado e amplitude conceitual, a curadoria digital difere da curadoria tal qual vem sendo compreendida ao longo do tempo. O que se observa, entretanto, é que a curadoria digital mostra alguma continuidade com as práticas tradicionais de curadoria. Independentemente de uma coleção ser constituída de objetos físicos ou digitais – ou seja, de átomos e moléculas ou de bits e bytes –, um curador deve: avaliar seu valor e relevância para a comunidade de usuários reais e potenciais; determinar a necessidade de preservação; documentar a origem e autenticidade; descrever, registrar e catalogar seu conteúdo; providenciar armazenamento e preservação em longo prazo; e proporcionar um meio de acesso e uso para os conteúdos (NATIONAL RESEARCH COUNCIL, 2015).

Quando voltamos a atenção para as coleções de dados de pesquisa, verificamos que os museus de história natural têm se dedicado há muito tempo à curadoria de dados que se manifestam na forma de coleções de espécimes físicas. O papel e as responsabilidades curatoriais dessas instituições ajudam a explicar como o termo **curadoria** está sendo aplicado também às concepções correntes de gestão e preservação de dados de pesquisa.

Antes das práticas acadêmicas se deslocarem para um reino digital ou para um paradigma do *big data*, os museus de história natural já tinham ampliado o seu conceito de curadoria antecipando a demanda por gestão e aprimoramento dos dados digitais (PALMER et al., 2013, p. 2).

Não obstante os termos **curadoria** e, mais especificamente, **curadoria digital** serem aplicados de muitas maneiras e em muitas áreas, há uma convergência de ideias em torno do que seja **curadoria de dados**. A definição adotada pelo RESEARCH DATA CANADA (2014) sintetiza com clareza o conceito: “atividade de gestão e promoção do uso dos dados desde o ponto da criação para assegurar que o dado esteja pronto para os propósitos correntes [e futuros] e esteja disponível para descoberta e reuso”. A curadoria de dados, nesse sentido, implica numa gestão dinâmica, que incorpora uma vasta gama de atividades tanto intelectuais e gerenciais como técnicas, que estão muito além da simples coleta e armazenamento de dados e informação.

A curadoria de dados se desdobra em ações planejadas, sistemáticas, intencionais e direcionadas que tornam os dados adequados para um propósito; se refere também a atividades que garantam que os dados permaneçam recuperáveis, acessíveis e utilizáveis enquanto as comunidades-alvo tiverem necessidade ou o direito de usá-los. Para coleções de dados dinâmicas – ou seja, que podem

variar ao longo do tempo -, a curadoria de dados pode significar a necessidade de enriquecimentos contínuos dos dados ou de atualizações para manter os dados adequados às suas finalidades.

Adicionar valor aos dados digitais para uso atual e futuro é uma parte essencial da curadoria de dados, visto que a maior parte dos dados não é naturalmente útil, nem imediatamente valiosa no momento em que é criada ou coletada. Os processos de curadoria é que reduzem ou eliminam ruídos presentes nos dados, e que detectam e corrigem erros e outras anomalias que possibilitam o aumento da utilidade imediata dos dados. Transformar dados digitais em informações úteis geralmente requer intervenção ativa por pessoas qualificadas e aplicações de software, o que significa, por exemplo, assinalar metadados descritivos e disciplinares, selecionar e avaliar, e integrar e processar dados por meio de ferramentas computacionais adequadas.

Altos níveis de curadoria envolvem atuações incisivas sobre os dados, como links para anotações e para outros materiais publicados (AUSTIN et al., 2015); agregações e visualização de dados são exemplos de serviços que se realizam por intervenção de pessoas e sistemas sobre as coleções de dados, com o propósito de adicionar valor a essas coleções e aumentar o seu potencial de reuso. É preciso lembrar que as informações de representação que descrevem o que os dados são e como interpretá-los são essenciais para os processos de curadoria e devem ser também objetos da curadoria de dados (JOHNSTON, 2017).

O princípio de usabilidade atual e futuro para os dados de pesquisa tem implicações importantes para a curadoria de dados. A ampliação da gama de usos dos dados por vários domínios disciplinares requer uma curadoria orientada para a diversidade contemporânea de usuários e de metodologias; por outro lado, o potencial de usabilidade no futuro, tanto dentro como fora do contexto em que foi criado ou recolhido pela primeira vez, coloca exigências adicionais sobre a curadoria, que implica em atenção às mudanças e atualizações das tecnologias subjacentes – hardware, software – e metadados, tanto para a preservação da informação digital como para a manutenção do acesso a ela (NATIONAL RESEARCH COUNCIL, 2015).

Além disso, como a informação digital é frágil, corruptível, facilmente alterada e sujeita a deleção acidental e intencional, manter a integridade da informação é um aspecto crítico da curadoria digital. Nesse sentido, a curadoria pode melhorar a integridade da informação digital e aumentar sua confiabilidade mediante procedimentos de segurança e do acesso restrito aos sistemas de curadoria, bem como de replicação, documentação de todas as transformações passadas pelos dados e por processos e procedimentos auditáveis (NATIONAL RESEARCH COUNCIL, 2015).

A curadoria de dados implica numa gestão comprometida que atua continuamente em todo ciclo de vida dos dados, iniciando-se na fase de planejamento dos dados. Esse fato tem um impacto importante sobre os fluxos tradicionais dos sistemas de informação, principalmente para as bibliotecas científicas, que tradicionalmente atuam com maior intensidade na gestão da pós-publicação, cujo objetivo é a disseminação de produtos finais, como livros, periódicos e teses. Apesar das dificuldades conceituais e práticas, da falta de profissionais especializados e da distância dos laboratórios e de seus pesquisadores, as bibliotecas de pesquisa se deslocam progressivamente para serem responsáveis pelas funções de gestão dos dados em suas respectivas instituições. Os processos do ciclo de curadoria de dados de pesquisa estão sendo integrados paulatinamente e de forma contextualizada nos espaços dos serviços de gestão de dados de pesquisa, tendo como suporte fundamental uma forte sobreposição com os sistemas de repositórios locais e os localizados remotamente (JOHNSTON, 2017).

O desafio dos curadores de dados é aplicar os princípios arquivísticos da biblioteconomia e da ciência da informação ao amplo espectro de objetos digitais complexos provenientes de várias disciplinas, e prepará-los para a ingestão, acesso e preservação de longo prazo em ambientes que facilitem a descoberta e o acesso ao mesmo tempo em que reforcem seus contextos, autenticidade e valor (JOHNSTON, 2017). Para tal, a curadoria de dados não pode prescindir de infraestruturas tecnológicas que viabilizem e tornem mais eficientes a colaboração entre os pesquisadores, reduzindo as restrições de velocidade, custo e distância entre os laboratórios e seus cientistas, aumentando o nível de atividades de pesquisa interinstitucionais e interdisciplinares e a taxa de coautoria.

## **INFRAESTRUTURA PARA A GESTÃO DE DADOS DE PESQUISA**

Cada campo disciplinar, embora em diferentes proporções, produz artigos de periódicos e de conferências, teses e dissertações, *preprints*, livros e outros materiais; as bibliotecas, por sua vez, selecionam, coletam, organizam e tornam disponível essa variedade de publicações provenientes de todas as áreas, criando uma continuidade cíclica que se estende por todo o universo científico. Não existe, entretanto, uma infraestrutura comparável voltada para as coleções de dados de pesquisa. Somente umas poucas áreas dispõem de mecanismos consolidados para publicação de dados em repositórios ou centros de dados; e algumas dessas áreas estão ainda no estágio de desenvolvimento de padrões e práticas que tornem seus dados mais acessíveis e preservados (BORGMAN, 2007).

A ampla base instalada de repositórios institucionais, que tem uma amplitude global, parece não ser perfeitamente adequada à gestão de dados de pesquisa em termos de concepção, modelo de dados e de fluxo de trabalho. Isso porque os repositórios institucionais foram projetados, desde sua origem, com o objetivo de arquivar e disponibilizar publicações acadêmicas; nos seus fluxos não há a preocupação em capturar e dar suporte às fases iniciais do trabalho de elaboração de produtos de pesquisa, sejam eles publicações ou coleções de dados. Considerando o longo e complexo ciclo de vida dos dados e a necessidade de capturá-los em diferentes estágios de produção e processamento, o modo praticado pelos repositórios institucionais, voltado prioritariamente para a publicação final de *e-prints*, pode ser inadequado para a gestão e curadoria de dados de pesquisa (HITCHCOCK, 2012). A manutenção das coleções de dados demanda conhecimento científico das áreas onde foram gerados e conhecimento tecnológico avançado para armazenar e organizar os dados para que eles sejam apropriadamente preservados e para que outros pesquisadores possam efetivamente consultar as informações que eles encerram por longo prazo (NIELSEN; HJORLAND, 2014).

Fica claro que, para que o ciclo tradicional de comunicação científica se ajuste mais rapidamente às novas imposições de compartilhamento de dados estabelecidos pelos vários *stakeholders* – agências governamentais de fomento, editores científicos, ciência aberta, para citar alguns –, é necessário a instalação de infraestruturas tecnológicas e gerenciais robustas que assegurem o depósito, arquivamento, preservação e acesso das coleções de dados de valor contínuo. A ótica que consubstancia as ações em torno do desenvolvimento dessas infraestruturas tem como princípio facilitar a pesquisa – e também o ensino – de forma distribuída, colaborativa e interdisciplinar, em domínios cujas metodologias estão baseadas em grandes volumes de recursos digitais (BORGMAN, 2007). Dados de pesquisa encerram uma gama extraordinária de informações, e os repositórios de dados são pontos centrais no estabelecimento de ecossistema de dados que viabilizem as dinâmicas necessárias à explicitação dessas informações, por meio da gestão e da curadoria.

## **REPOSITÓRIO DE DADOS DE PESQUISA: INFRAESTRUTURA ESSENCIAL**

O requisito essencial, em termos de infraestrutura tecnológica e gerencial e de sustentabilidade econômica e política de longo prazo, para que um serviço de curadoria de dados se efetive, é que as coleções de dados estejam gerenciadas e disponíveis por meio de um sistema de repositório digital de dados de pesquisa. “Atualmente os repositórios despontam como o ponto central de um moderno

ecossistema de dados que se desenvolve impulsionado pelas exigências emergentes de dados de pesquisa disseminados de forma ética e aberta” (BACKER; DUERR, 2017, p. 6), tendo como pressuposto básico o acesso instantâneo a coleções de dados de pesquisa por comunidades de pesquisadores dispersos globalmente desejosos de compartilhar, compreender, agregar e sintetizar resultados de pesquisa (UZWYSHYN, 2016). Os repositórios de dados, dessa forma, se configuram como um sistema de informação, baseado em uma plataforma tecnológica e organizacional, que apoia os pesquisadores no armazenamento, arquivamento, preservação de suas próprias coleções de dados; e, ao mesmo tempo, no descobrimento e acesso a coleções de dados de outras fontes (KINDLING et al., 2017).

A concepção dos repositórios de dados foi moldada e instruída pelo esforço desenvolvido por muitas áreas – como computação, arquivologia, biblioteconomia e ciência da informação e ainda áreas disciplinares específicas – que precisam articular conhecimentos e práticas para lidar com ambientes, papéis, culturas e tecnologias em constante transição. O objetivo é gerenciar mudanças para assegurar que a estabilidade das coleções de dados seja mantida com requisitos apropriados de integridade, autenticidade e proveniência, e que elas possam ser recuperadas, acessadas e interpretadas tendo como perspectiva a possibilidade de reuso de seus conteúdos por longo prazo. O cientista da computação e de dados Jim Gray, por exemplo, já no começo do século XXI, trabalhava nessa direção: uma vez publicado em um repositório e apropriadamente documentado, os dados de pesquisa podem se manter disponíveis para sempre, de forma que outros cientistas possam reproduzir os resultados e realizar novos experimentos a partir desses dados. Dessa forma, os dados podem ser usados muito tempo depois que o projeto de pesquisa que os gerou tenha finalizado (GRAY et al., 2002).

Os repositórios de dados foram inicialmente projetados como centros de dados disciplinares voltados para áreas altamente estruturadas e que produzem grandes quantidades de dados, como física, genética, astrofísica e ciências ambientais, que resultaram do desenvolvimento de infraestruturas nacionais e internacionais de grandes proporções como o Protein Data Bank<sup>1</sup>. Ultimamente, a demanda por repositórios de dados está emergindo no contexto da cauda longa da ciência (*long tail of science*), ou seja, no contexto dos domínios científicos nos quais atividades científicas são desenvolvidas num grande número de laboratórios relativamente pequenos e por pesquisadores individuais que coletivamente produzem a maioria dos resultados científicos (ASSANTE et al., 2016). Nesses ambientes, cuja pesquisa produzida é específica, inovadora e de ponta, se reconhece que faltam repositórios temáticos talhados às coleções de dados produzidas; e que, para publicar os seus dados, os pesquisadores se utilizam de

1 Disponível em: <[158](http://www ww p d b . o r g / > . Acesso em: 18 jul. 2017.</a></p></div><div data-bbox=)

repositórios multidisciplinares ou soluções inadequadas, como publicar diretamente em suas páginas web.

Os repositórios são desenvolvidos com objetivos, modelos de dados, tecnologias e padrões, produtos e serviços, fluxo de trabalho, ciclos temporais e infraestruturas que atendam a uma determinada comunidade científica, disciplina, projeto ou tipo de dado específico. Há inúmeras classificações possíveis para essa grande variedade de repositórios; entretanto, a que diferencia os repositórios multidisciplinares dos repositórios disciplinares ou temáticos e, dentre estes, os que se especializam em dados homogêneos, parece ser a mais eloquente para a presente análise.

As plataformas disciplinares se voltam para domínios específicos como astronomia e física de partículas ou, mais especificamente, para tipos particulares de dados, como, por exemplo, a base de dados de biomodelos do *European Bioinformatics Institute*<sup>2</sup> (EBI). Em geral possuem modelos de dados adequados à representação das coleções de dados e oferecem uma carteira de serviços mais orientados às especificidades da área. Plataformas especializadas em aceitar dados homogêneos são capazes de projetar sistemas voltados para organizar, preservar e disseminar um tipo particular de dados, podendo, por conseguinte, oferecer serviços mais robustos e avançados para os dados, como: visualização, subcoleções, agregações, reprojeções, agregações e análises *on-the-fly*.

Os repositórios multidisciplinares, por sua vez, gerenciam coleções de dados de diversas áreas, estruturas e tipos, que implica em representação limitada e serviços básicos. Entretanto, a demanda por esse tipo de repositório está crescendo rapidamente devido às pressões impostas aos pesquisadores por vários atores (governo, agências de financiamento, editores, entre outros) por gestão e compartilhamento sistemático dos dados (ASSANTE et al., 2016); exemplo típico é o Dryad<sup>3</sup>, repositório de dados recomendado pelos editores de periódicos científicos para depósito e publicação de dados subjacentes aos artigos. Essas plataformas são essencialmente serviços de armazenamento e compartilhamento, e não repositórios de preservação de longo prazo.

Além de oferecer uma base tecnológica para a execução dos processos de gestão de longo prazo das coleções de dados, os repositórios se configuram também como um componente importante das infraestruturas de informação acadêmica, com influência cada vez mais contundente sobre o ciclo tradicional de comunicação científica. Nessa direção, os repositórios têm um papel importante nas interações que envolvem a validação do trabalho de pesquisa e na própria dinâmica social da comunicação científica. A possibilidade de se ter os dados de

2 Disponível em: <<https://www.ebi.ac.uk/biomodels-main/>>. Acesso em: 18 jul. 2017.

3 Disponível em: <<https://www.datadryad.org/>>. Acesso em: 18 jul. 2017.

pesquisa disponíveis *on-line*, indexados, documentados e anotados relativos a uma pesquisa publicada ou pré-publicada num artigo acadêmico, redimensiona a revisão por pares, estendendo-a a uma comunidade mais ampla e conectada em rede. “Um repositório permite exame, prova, revisão, transparência de resultados de pesquisa por outros especialistas que vão além da revisão por pares do artigo acadêmico publicado” (UZWYSHYN, 2016, p. 1). Os repositórios de dados também permitem a publicação de resultados negativos, que antes ficavam ocultos, provenientes de experimentos que não deram certo. Isso permite que outros pesquisadores evitem os caminhos sem volta tentados anteriormente (UZWYSHYN, 2016). Permeados por uma política formalmente explicitada, os repositórios apoiam os pesquisadores no intercâmbio de dados de forma que as exigências de propriedade intelectual, privacidade, ética e tempo de embargo sejam observadas. Todas essas características somadas conferem um papel importante aos repositórios em relação a alguns pressupostos fundamentais da ciência, como os princípios da reprodutibilidade e autocorreção.

Com a rápida inserção dos repositórios de dados nas bases mundiais de informação para a pesquisa, as organizações responsáveis por essas plataformas estão cada vez mais interessadas em avaliar suas infraestruturas e serviços ancorados em arcabouços de avaliação, auditoria, certificação e autoavaliação que sejam reconhecidos pelas comunidades científicas dos seus respectivos domínios. A certificação é importante porque promove a confiança na usabilidade e persistência dos dados que estão sendo compartilhados. Os processos de certificação e autoavaliação aplicados aos repositórios têm como documento primordial o Modelo de Referência Open Archival Information System (OAIS) (CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEM, 2012), que influenciou o desenvolvimento de diferentes metodologias de certificação.

Os repositórios, por meio de vários tipos de protocolos e tecnologias, com destaque para o Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)<sup>4</sup>, possibilitam o intercâmbio de informações – dados e metadados – com outras plataformas de forma harmônica e integrada e com propósitos específicos. “A interoperabilidade, em particular, desempenha um papel importante na garantia de que os dados podem ser encontrados e reutilizados por pesquisadores e instituições fora do grupo de pesquisa original” (AMORIM et al., 2015, p. 2).

Conforme o modelo proposto por Mayernik et al. (2013), as dinâmicas no âmbito de um repositório são representadas por meio de um conjunto de serviços de gestão de dados que se configuram em quatro camadas, desde uma camada mais física até uma camada mais abstrata, que é definida pelos serviços de curadoria digital. Nesse sentido, temos:

4 Disponível em: <<https://www.openarchives.org/pmh/>>. Acesso em: 18 jul. 2017.



- a) armazenamento, como a camada de serviço mais física, tratando de bits e mídias, backups e segurança da informação;
- b) arquivamento, tratando da proteção dos dados, incluindo integridade, fixidez e identificadores;
- c) preservação, assegurando que os dados possam ser plenamente usados e interpretados por uma comunidade-alvo por longo prazo;
- d) na camada mais abstrata a curadoria, que adiciona valor às coleções de dados por todo o ciclo de vida, tendo como perspectiva o reuso.

Os repositórios de dados de pesquisa têm como objetivo fundacional garantir o acesso contínuo e aberto – agora e no futuro – aos resultados de pesquisa que se manifestam na forma de dados, e que são considerados parte importante do patrimônio digital da humanidade. As ações dos repositórios de dados nessa direção trazem benefícios relevantes para o mundo da pesquisa científica e para toda a sociedade. Partindo da análise preliminar de Sales (2014), a seguir são apresentados alguns dos benefícios mais perceptíveis:

- a) visibilidade dos dados: amplia a visibilidade dos dados de pesquisa, tornando-os disponíveis via sistemas web, permitindo que eles sejam localizados, recuperados, acessados, reusados e citados mais frequentemente (geralmente só é disseminada pelos canais formais a fração de dados – cerca de 10% – que está registrada nos artigos publicados);
- b) compartilhamento de dados: os repositórios, pela sua capacidade de agregação e de organização de recursos informacionais dispersos no tempo e no espaço, e como instrumentos de socialização de comunidades e grupos pesquisadores ao redor desses recursos, tornam-se um dispositivo importante de troca de experiências e compartilhamento de dados;
- c) crédito ao autor dos dados: os repositórios de dados tornam possível identificar as coleções de dados e seus autores de forma unívoca e persistente, permitindo que os autores sejam reconhecidos, citados, avaliados e recompensados pelo trabalho intelectual de coleta, geração e organização dos dados;
- d) preservação digital: oferece um ambiente tecnológico, gerencial e de padronização propício para a preservação de longo prazo dos dados de pesquisa de valor contínuo, especialmente para os dados observacionais;
- e) memória científica e transparência: contribui para a formação da memória científica das instituições no que diz respeito aos dados, complementando os repositórios institucionais que estão focados nas publicações acadêmicas; na qualidade de registro das atividades de pesquisa das instituições, contribui também com os princípios de transparência, tão em voga nos tempos atuais;

- f) segurança dos dados: oferece sistema de armazenamento seguro, esquemas de backup e segurança física que se contrapõem ao armazenamento informal em mídias portáteis e computadores pessoais frequentemente usados pelos pesquisadores;
- g) disponibilidade: permite que os dados estejam disponíveis on-line para serem acessados, baixados, visualizados e processados por pessoas ou por sistemas;
- h) curadoria digital: proporciona um ambiente apropriado para os processos de avaliação, de adição de valor, reformatação, agregação e recriação de dados promovidos pela curadoria digital;
- i) serviços inovadores: abre possibilidades de criação de novos serviços de informação para pesquisadores, gestores e financiadores de pesquisa a partir da análise e integração dos dados arquivados com fontes internas e externas à instituição;
- j) reuso dos dados: aumenta o grau de reuso e reinterpretação dos dados, possibilitando a realização de novas pesquisas de caráter interdisciplinar; minimiza a duplicação de esforços e otimiza os investimentos na coleta e geração de dados;
- k) redes de repositórios: permite, por meio de protocolos de interoperabilidade, como o OAI-PMH, a formação de redes de repositório de dados; abre a possibilidade de inserção dos repositórios de dados às redes interoperáveis definidas pelo padrão *Linked Data*;
- l) indicador de qualidade e produtividade da instituição: as coleções de dados organizados e arquivados no repositório são evidências da qualidade e da relevância das atividades de pesquisa da instituição, atestando a sua produtividade e seu valor acadêmico.

## CONSIDERAÇÕES FINAIS

A falta de infraestruturas centradas em repositórios digitais para a gestão de dados de pesquisa provoca uma descontinuidade no fluxo de comunicação científica e na integração dos diversos sistemas que subsidiam a publicação de produtos de pesquisa como artigos, teses e dados. O que se observa é que, fora do *big science*, grande parte das coleções de dados existentes é gerenciada por pesquisadores individualmente, por programas ou laboratórios dentro dos limites disciplinares onde foram geradas como um bem particular, orientadas para o uso imediato e próximo e sem olhar para o futuro. Estes ativos informacionais estão longe das bibliotecas e arquivos e das estruturas de publicação acadêmica e sem potencial de compartilhamento e reuso e sem perspectiva de preservação. E isso não se dá somente por falta de infraestruturas tecnológicas e gerenciais

adequadas. Amplificando essa descontinuidade está o fato de que o compartilhamento de dados não é um consenso no mundo da ciência; embora algumas disciplinas tenham uma longa tradição no compartilhamento de dados, um número grande de pesquisadores reluta em fazê-lo, mesmo quando os projetos de pesquisa são financiados com recursos públicos; mesmo quando o resultado mais eloquente do compartilhamento de dados é o aumento das atividades de pesquisa interinstitucionais e interdisciplinares, que se refletem na taxa de publicações em coautoria, citações e de projetos colaborativos. Isso indica que há ainda um longo caminho a percorrer, que deve estar paralelo ao desenvolvimento e implantação de sistemas de repositórios e da integração desses sistemas às infraestruturas nacionais e globais de informação para a pesquisa: políticas nacionais e institucionais de dados de pesquisa; esquema de recompensa para os pesquisadores que organizam seus dados; fóruns interinstitucionais para a discussão dos desafios a serem enfrentados; financiamento contínuo, legislação, pesquisa, cursos, manuais e guias, treinamentos para profissionais de informação, de computação e, o mais importante, para os pesquisadores que têm responsabilidades sobre os dados que eles coletam e geram.

## REFERÊNCIAS

ABBOTT, D. **What is digital curation?** Edinburgh: Digital Curation Centre, 2008. Disponível em: <<http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/what-digital-curation>>. Acesso em: 01 maio 2017.

AMORIM, R. C. et al. A comparative study of platforms for research data management: interoperability, metadata capabilities and integration potential. In: ROCHA, A. et al. (Ed.). **New contributions in information systems and technologies**. Heidelberg: Springer, 2015. p. 101-111. Disponível em: <[https://link.springer.com/chapter/10.1007/978-3-319-16486-1\\_10](https://link.springer.com/chapter/10.1007/978-3-319-16486-1_10)>. Acesso em: 25 maio 2017.

ASSANTE, M. et al. Are scientific data repositories coping with research data publishing? **Data Science Journal**, v. 15, 2016. Disponível em: <<http://datascience.codata.org/article/10.5334/dsj-2016-006/>>. Acesso em: 01 maio 2017.

AUSTIN, C. et al. Research data repositories: review of current features, gap analysis, and recommendations for minimum requirements. **IASIST Quarterly**, winter 2015. Disponível em: <[http://www.iassistdata.org/sites/default/files/vol\\_39\\_4\\_austin.pdf](http://www.iassistdata.org/sites/default/files/vol_39_4_austin.pdf)>. Acesso em: 01 maio 2017.

BACKER S. K.; DUERR E. R. Research and changing nature of data repositories. In: JOHNSTON, L. (Ed.). **Curating research data: practical strategies for your digital repository**. Illinois: Association of College and Research Libraries, 2017. p. 33-60. Disponível em: <[http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/booksanddigitalresources/digital/9780838988596\\_crd\\_v1\\_OA.pdf](http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/booksanddigitalresources/digital/9780838988596_crd_v1_OA.pdf)>. Acesso em: 01 maio 2017.

BORGMAN, C. L. **Scholarship in the digital age: information, infrastructure and the internet**. Cambridge: The MIT Press, 2007.

BOULTON, G. et al. **Science as an open enterprise**. London: The Royal Society, 2012. Disponível em: <[https://royalsociety.org/~media/Royal\\_Society\\_Content/policy/projects/sape/2012-06-20-SAOE.pdf](https://royalsociety.org/~media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf)>. Acesso em: 01 maio 2017.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEM. **Reference Model for an Open Archival Information System (OAIS)**: recommended practice CCSDS 650.0-M-2. Washington: CCSDS, 2012. Disponível em: <<https://public.ccsds.org/pubs/650x0m2.pdf>>. Acesso em: 01 maio 2017.

GRAY, J. et al. **Online scientific data curation, publication, and archiving**. Redmond: Microsoft Research Corporation, 2002. Disponível em: <<https://pdfs.semanticscholar.org/8cf6/ce243af68ceeda6ffc2b1e59ba5d1264920.pdf>>. Acesso em: 01 maio 2017.

HITCHCOCK, S. **What can research data repositories learn from open accesses?** Part 1. 2012. Disponível em: <<http://datapool.soton.ac.uk/2012/05/24/what-can-research-data-repositories-learn-from-open-access-part-1/>>. Acesso em: 01 maio 2017.

JOHNSTON, L. R. Introduction to Data Curation. In: JHONSTON, L. R. (Ed.). **Curating research data**: practical strategies for your digital repository. Illinois: Association of College and Research Libraries, 2017. p. 1-29. Disponível em: <[http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/booksanddigitalresources/digital/9780838988596\\_crd\\_v1\\_OA.pdf](http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/booksanddigitalresources/digital/9780838988596_crd_v1_OA.pdf)>. Acesso em: 01 maio 2017.

KINDLING, M. et al. The landscape of research data repositories in 2015: a re3data analysis. **D-Lib Magazine**, v. 23, n. 3/4, mar./apr. 2017. Disponível em: <<http://www.dlib.org/dlib/march17/kindling/03kindling.html>>. Acesso em: 01 maio 2017.

KINDLING, M.; SCHIRMBACHER, P. Die digitale Forschungswelt als Gegenstand der Forschung. **Information - Wissenschaft & Praxis**, v. 64, n. 2/3, p. 127-136, apr. 2013. Disponível em: <<https://edoc.hu-berlin.de/handle/18452/14320>>. Acesso em: 04 aug. 2017.

LEE, C.; TIBBO, H. Digital curation and trusted repositories: steps toward success. **Journal of Digital Information**, v. 8, n. 2, 2007. Disponível em: <<https://journals.tdl.org/jodi/index.php/jodi/article/view/229/183>>. Acesso em: 01 maio 2017.

LEIDEN UNIVERSITY. **Data management**. 2015. Disponível em: <<https://www.library.universiteitleiden.nl/research-and-publishing/datamanagement>>. Acesso em: 15 maio 2016.

LIU, S. et al. A survey on information visualization: recent advances and challenges. **Vis Comput**, v. 30, p. 1373-1393, 2014. Disponível em: <<http://www.shixialiu.com/publications/InfovisSurvey/paper.pdf>>. Acesso em: 01 maio 2017.

LORD, P.; MACDONALD, A. e-Science data curation. **JISC**, nov. 2004. Disponível em: <<http://webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/media/documents/publications/esciencedatacuration.pdf>>. Acesso em: 01 maio 2017.

MARTINEZ-URIBE, L.; MACDONALD, S. A new role for the academic librarian: data curation. **Profesional de la Información**, v. 17, n. 3, p. 273-280, 2008.

MAYERNIK, M. S. et al. Data conservancy provenance, context, and lineage services: key components for data preservation and curation. **Data Science Journal**, v. 12, 2013. Disponível em: <<http://datascience.codata.org/articles/abstract/10.2481/dsj.12-039/>>. Acesso em: 25 maio 2017.

MAYERNIK, M. S. et al. Linking publication and data: challenges, trends and opportunities. **D-Lib Magazine**, v. 22, n. 5/6, may/jun. 2016. Disponível em: <<http://www.dlib.org/dlib/may16/mayernik/05mayernik.html>>. Acesso em: 01 maio 2017.

NATIONAL RESEARCH COUNCIL. **Preparing the workforce for digital curation**. Washington: National Academic Press, 2015. Disponível em: <<https://www.nap.edu/catalog/18590/preparing-the-workforce-for-digital-curation>>. Acesso em: 01 maio 2017.

NATIONAL SCIENCE BOARD. **Long-lived digital data collections**: enabling research and education in the 21st century. Virginia: National Science Foundation, 2005. Disponível em: <<http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>>. Acesso em: 01 maio 2017.

NIELSEN, H. J.; HJØRLAND, B. Curating research data: the potential roles of libraries and information professional. **Journal of Documentation**, v. 70, n. 2, p. 221-240, 2014.

PALMER, C. et al. Foundation of data curation: the pedagogy and practice of “Purposeful Work” with research data. **Archival Journal**, n. 3, 2013. Disponível em: <<http://www.archivejournal.net/issue/3/archives-remixed/foundations-of-data-curation-the-pedagogy-and-practice-of-purposeful-work-with-research-data/>>. Acesso em: 01 maio 2017.

RESEARCH DATA CANADA. **Glossary of terms and definitions**. 2014. Disponível em: <<http://www.rdc-drc.ca/glossary>>. Acesso em: 09 ago. 2017.

SALES, L. F. **Integração semântica de publicações científicas e dados de pesquisa**: proposta de modelo de publicação ampliada para a área de Ciências Nucleares. 2014. 268 f. Tese (Doutorado em Ciência da Informação) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014. Disponível em: <<http://repositorio.ibict.br/bitstream/123456789/874/1/LUANA%20SALES%20D.pdf>>. Acesso em: 25 maio 2017.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN, 2015. Disponível em: <[http://carpedien.ien.gov.br:8080/bitstream/ien/1624/1/GUIA\\_DE\\_DADOS\\_DE\\_PESQUISA.pdf](http://carpedien.ien.gov.br:8080/bitstream/ien/1624/1/GUIA_DE_DADOS_DE_PESQUISA.pdf)>. Acesso em: 15 maio 2017.

UZWYSHYN, R. Research data repositories: the what, when, why and how. **Computers in libraries**, v. 36, n. 3, apr. 2016. Disponível em: <<http://www.infotoday.com/cilmag/apr16/Uzwyshyn--Research-Data-Repositories.shtml>>. Acesso em: 01 maio 2017.

VALLE, M. **Scientific data management**: an introduction. 2017. Disponível em: <<http://mariovalle.name/sdm/scientific-data-management.html>>. Acesso em: 01 maio 2017.



# ADOÇÃO DE PADRÕES DE METADADOS PARA REPOSITÓRIOS DE DADOS DIGITAIS NA CIÊNCIA ABERTA

Adriana Carla Silva de Oliveira









## A CIÊNCIA ABERTA E A PRÁXIS DA E-SCIENCE PARA DADOS DE PESQUISA

A comunicação científica está evoluindo para a promoção de uma ciência aberta e inovadora apoiada no quarto paradigma da ciência e em novas práticas científicas. Ciência aberta é a tradução da expressão *open science* que surgiu no século XVI após a revolução pós-renascentista, a partir da propriedade dos bens intelectuais e com o surgimento da prensa móvel, ressurgindo nas últimas décadas como um grande esquema do movimento de *open access* (FROSIO, 2014). Muito embora não seja um termo recente para o contexto atual, a nova acepção da ciência aberta traz um olhar diferente para a ciência contemporânea a partir do advento dos movimentos de *open access*, e mais recentemente, *open science* e *data science*.

Assim, o capítulo se reporta a ciência aberta como um termo amplo que envolve várias dimensões. A multidimensionalidade é constituída por dimensões que abordam aspectos epistemológicos, teóricos, políticos, éticos, legais, culturais, sociais, morfológicos, técnicos e tecnológicos que permeiam o processo de investigação científica orientado aos dados de pesquisa. Inclui, também, o acesso livre às “publicações científicas, dados científicos abertos, ferramentas científicas abertas, *hardware* científico aberto, cadernos científicos abertos, *wikipesquisa*, ciência cidadã e educação aberta” (ALBAGLI, 2015, p. 15), como produtos de uma ciência colaborativa, compartilhada, aberta e comum.

No contexto da comunicação científica, a ciência aberta conduz um novo fazer científico - que pressupõe regras, modelos e práticas distintas da ciência tradicional - constituído por dados de pesquisa. Ademais, conduz novas investigações e hipóteses para o uso, reuso e reprodutibilidade desses dados, sob a égide do quarto paradigma, o qual fomenta o surgimento de um fazer científico baseado no compartilhamento e na colaboração.

Especificamente, o quarto paradigma se traduz pelo uso intensivo de dados primários como produtos preliminares do processo de investigação e pela transição de práticas científicas do modelo tradicional para a nova abordagem nomeada de *práxis* da *e-Science*. Reflete, também, a capacidade de inferir significado aos dados de pesquisa e direcionar para um novo sentido. Dar sentido e gerar novos significados com o uso e reuso de dados é uma das facetas desse quarto

paradigma. Importante salientar que o surgimento de novo paradigma é sempre revolucionário no contexto da comunicação científica, pois provoca mudanças significativas, à medida que altera as formas de produção, gestão e acesso das investigações científicas contemporâneas (BORGMAN, 2010; OLIVEIRA et al., 2015).

No bojo da ciência, a comunicação científica requer inovação, não só no entendimento das descobertas científicas, mas no seu processo operacional de mudança do *modus operandi*, pois a:

inovação nunca acontece em um vácuo; inovação requer comunicação. [...] realizar todos os estágios de pesquisa, desenvolvimento, desenho, e produção tão eficientes e efetivos quanto possíveis, são necessários para que a informação possa ser clara e entendida (TENOPIR; KING, 2004, p. 11, tradução nossa).

Então, a essência dessa *práxis* está na mudança do *modus operandi* de fazer, divulgar, acessar e compartilhar dados de maneira que possibilite o compartilhamento, a interatividade, a convergência e a colaboração, entre pesquisadores e instituições, para o uso, o reuso e a reprodutibilidade dos mesmos. Constitui, no contexto científico contemporâneo, a real essência de uma pesquisa compartilhada, colaborativa, aberta e inovadora.

Para tanto, a *práxis* da *e-Science* se apoia em três eixos:

- a) adotar o modelo de ciclo de vida que garanta o efetivo gerenciamento dos dados primários;
- b) proporcionar a ciberinfraestrutura sustentável para uso, reuso e reprodutibilidade dos dados primários em longo prazo;
- c) possibilitar o compartilhamento e a colaboração científica entre diferentes pesquisadores, profissionais, instituições e países.

Tais eixos são essenciais para que o processo de pesquisa baseado em dados seja eficiente e sustentável em longo prazo.

Borgman (2003) previu uma nova era digital, a qual chamou de *Global Information Infrastructure* (GII), comparando o momento ao feito de Gutenberg. Afirma que “Gutenberg aperfeiçoou a impressão com tipo móvel. Foi a abordagem holística que o possibilitou aperfeiçoar cada fase do processo [...], contribuindo, portanto, significativamente para a tecnologia ou comunicação humana” (BORGMAN, 2007, p. 30, tradução nossa). A infraestrutura global de informação a que se refere a autora aponta para a reflexão na seara da pesquisa tradicional para uma nova concepção de pesquisa aberta. É a natureza colaborativa e compartilhada, associada à interdisciplinaridade entre as áreas que altera o *status quo* da ciência tradicional para a *e-Science*.

O termo *e-Science* ou *eScience* é definido como a ciência de computação intensiva, realizada em ambientes de rede altamente distribuídos que utilizam volumosos conjuntos de dados, os quais permitem a coleta, o processamento, a preservação, a análise e o armazenamento de grande quantidade de dados em diferentes disciplinas (GRAY, 2009). Representa, pois, as práticas do quarto paradigma científico. Sua origem remonta à Segunda Guerra Mundial, quando houve uma explosão de publicações científicas e o crescimento exponencial do conhecimento científico em disciplinas especializadas, emergindo o chamado fenômeno *big science*.

No século XXI, “a pesquisa colaborativa e aberta que produz grande volume de dados deu origem aos termos: *big data*, *cyberscience*, *cyber infrastructure* e *e-Science*” (COSTA; CUNHA, 2015, p. 128) para representar as inovações, descobertas e teorias no escopo da ciência contemporânea. Pesquisadores adeptos do processo de compartilhamento e colaboração científica estão focados na recompensa científica, reuso e reprodução de experimentos de pesquisa semelhantes, cujo intuito é evitar erros inúteis e aprimorar o processo de investigação (GRAY, 2009; AMBINDER; MARCONDES, 2012).

Além disso, “a relevância dos dados no contexto das ‘big Sciences’, como Astronomia, Física e Biologia, conduziu não somente ao surgimento de novos modelos de ciência - coletivamente chamados de ‘quarto paradigma científico’ ou ‘eScience’ [...]” (BORGMAN, 2010, p. 2, tradução nossa), mas aos modelos e práticas a partir da utilização e do valor que os dados de pesquisa passaram a ter em áreas consideradas de alta complexidade e especializadas das grandes ciências. Em áreas como “as ciências da vida e outros campos que dependem de dados observacionais, a situação é mais complexa, à medida que os dados são usados para modelagem avançada, a qual cria novos conjuntos de dados” (TENOPIR et al., 2011, p. 4, tradução nossa), requerendo políticas e infraestrutura sustentáveis para apoiar a investigação.

No cenário internacional, políticas institucionalizadas no âmbito das agências de financiamento, instituições de pesquisa e ensino, organizações governamentais e não governamentais emitiram:

nos últimos anos declarações e políticas ressaltando a necessidade para solicitação de preservação de dados, e algumas agências de financiamento começaram a exigir que os projetos de dados financiados, sejam depositados em arquivo público (INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH, 2012).

Projetos de pesquisas financiados com recursos públicos seguem recomendações de boas práticas e adotam ferramentas e modelos abertos para todo o

processo de gerenciamento de dados de pesquisa, a fim de torná-los acessíveis, interoperáveis e abertos.

Percebe-se que a comunidade científica anseia por novos modelos de publicações que expressem a complexidade e a dinâmica da pesquisa científica contemporânea (SALES; SAYÃO; SOUZA, 2013). Os novos modelos científicos possibilitam claramente a pesquisa desenvolvida, seus métodos, seu material, os dados e conhecimentos gerados, além de revelar aspectos intrínsecos da investigação científica, a exemplo da intenção do pesquisador e a avaliação dos resultados de forma mais eficaz e interativa.

No cenário nacional, a iniciativa brasileira na área de biodiversidade é um exemplo de gestão, integração e monitoramento de dados ecológicos. O projeto pioneiro com dados da biodiversidade no Brasil é desenvolvido pela Universidade de São Paulo, por instituições de pesquisas ambientais (nacionais e internacionais) e o Ministério do Meio Ambiente (MMA). Os dados foram adequados à sistemática e à infraestrutura tecnológica compartilhada com unidades de conservação ambiental e o Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio).

No que tange à operacionalização da *práxis* científica, ela é realizada pelo ciclo de vida dos dados. Procedimentos, padrões, protocolos, modelos, ferramentas tecnológicas e licenças públicas são configurados em termos de conformidade internacional dentro de uma ciberinfraestrutura sustentável.

Em um projeto de dados, inúmeros pesquisadores, colaboradores e instituições são partícipes, de modo a garantir procedimentos internacionais confiáveis, qualitativos e sustentáveis, aumentando a credibilidade e veracidade da pesquisa, além de evitar o uso inapropriado de termos ou parâmetros.

A modelagem da pesquisa baseada no ciclo de vida dos dados constitui o principal eixo de condução da *práxis* científica da *e-Science*.

## **MODELAGEM DO CICLO DE VIDA E CIBERINFRAESTRUTURA SUSTENTÁVEL**

Como já mencionado, a *práxis* da *e-Science* é apoiada por uma infraestrutura tecnológica, metodológica e técnica denominada de ciberinfraestrutura que suporta o ambiente aberto e compartilhado da pesquisa, como um dos componentes de sustentabilidade. O processo de gestão dos dados é modelado em estágios que garantem o planejamento, gerenciamento, curadoria e preservação dos dados em todas as etapas da investigação.

A ciberinfraestrutura que permeia toda a *práxis* é sistematizada por aspectos metodológicos, tecnológicos e técnicos previstos no decorrer dos estágios do ciclo de vida. É a partir da adoção de um ciclo de vida que as ações de planejar, coletar, descrever, armazenar, preservar, analisar e visualizar os dados são garantidas e asseguradas, quanto à qualidade, a fidedignidade, a interoperabilidade e a curadoria por um longo prazo.

É por meio da ciberinfraestrutura que pesquisadores, curadores e colaboradores envolvidos no processo de pesquisa e de gerenciamento de dados, fazem uso da alta tecnologia, computação intensiva, ferramentas e recursos para lidar com dados (estruturados e desestruturados). Os recursos tecnológicos propiciam as condições para captura, armazenagem, gerenciamento, preservação, visualização e análise de dados de pesquisa, “tornando os sistemas de computação vitais para o moderno ambiente de pesquisa” (COSTA; CUNHA, 2015, p. 130). Ademais, a infraestrutura global proposta requer o estabelecimento de alianças institucionais e o desenvolvimento de políticas governamentais para que transformações possam ser realizadas.

O estudo de Atkins et al. (2003) aponta para uma ciberinfraestrutura constituída em duas camadas:

- a) a camada subjacente, composta pelos componentes integrados eletro-ópticos da computação, que favorecem a armazenagem e comunicação integradas;
- b) a camada superior, composta por programas de software, serviços, instrumentos, dados, informações, conhecimentos e práticas sociais aplicáveis aos projetos específicos, treinamento e comunidades de pesquisa.

Entre as duas camadas, a ciberinfraestrutura integra os **componentes internos** (algoritmos, hardware, software) e o **ambiente externo** (comunicação científica, pessoas e instituições envolvidas) no processo de pesquisa.

As camadas que constituem a ciberinfraestrutura “devem fornecer uma plataforma eficaz e eficiente para o empoderamento das comunidades específicas de pesquisadores e, eventualmente, inovar e revolucionar o que eles fazem, como eles fazem, e quem participa” (ATKINS et al., 2003, p. 5, tradução nossa). A garantia de padrão internacional e conformidade para a sustentabilidade do compartilhamento, colaboração e gestão do processo completo dos dados deve ser atendida.

Como forma de ilustrar teoricamente os estágios do modelo de ciclo de vida, o estudo traz a modelagem do *Data Lifecycle* do Projeto *DataONE* conforme Figura 1. O *Data Lifecycle* do *DataONE* é constituído por oito estágios que sistematiza as

ações: Planejar (*Plan*); Coletar (*Collect*); Assegurar (*Assure*); Descrever (*Describe*); Preservar (*Preserve*); Descobrir (*Discover*); Integrar (*Integrate*); e Analisar (*Analyze*).

A seguir, os estágios são descritos com algumas exemplificações de soluções tecnológicas e técnicas.

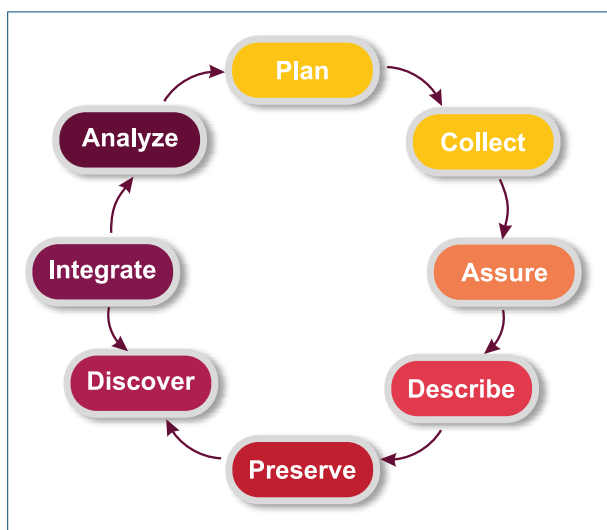


Figura 1 – Modelo de Ciclo de Vida dos Dados

Fonte: Oliveira (2016).

## PRIMEIRO ESTÁGIO DO CICLO: PLANEJAR (*PLAN*)

O estágio planejar, como o próprio nome diz, está relacionado ao planejamento, e descreve o mapeamento dos processos metodológicos e técnicos, bem como a ciberinfraestrutura tecnológica necessária para a execução de todos os estágios subsequentes do ciclo de vida.

Na execução desse estágio, a prática aconselhada é utilizar ferramentas tecnológicas para auxiliar pesquisadores e colaboradores, agências de fomento e instituições de ensino e pesquisa, a desenvolverem os planos com base em orientações, regras estabelecidas e uma metodologia que norteie todos os estágios do processo investigativo.

Exemplo é o *Data Management Plan Tool* (DMPTool), utilizado como ferramenta aberta e livre para o cumprimento dessa fase, fornecendo guia de boas práticas aos projetos orientados por dados de pesquisa. O intuito dessa ferramenta é favorecer a eficiente gestão de dados e proporcionar que os dados sejam usados, compartilhados e reutilizados.

O *Data Management Plan Tool* (DMPTool) é bastante utilizado por agências de fomento, instituições de ensino e científicas, bibliotecas e pesquisadores em geral que gerenciam seus dados por meio desta ferramenta. É um *open-source*, aplicativo gratuito, *on-line*, que auxilia na criação dos planos de gestão de dados. A ferramenta é baseada em cinco componentes: informações sobre os dados e formato de dados; conteúdo de metadados e padrões; políticas para acesso, compartilhamento e reuso; gerenciamento e armazenagem (longo prazo); e orçamento.

Recentemente, a ferramenta *Data Management Plan Online* (DMPonline) foi atualizada pelo *Digital Curation Centre* (DCC) do Reino Unido. A finalidade do DCC é ajudar equipes de pesquisa a cumprir as exigências das agências de fomento e expectativas institucionais no escopo da ciência aberta (DIGITAL CURATION CENTRE, 2016). Estes planos ou DMPs fazem parte de uma política mandatória adotada pelas agências de financiamento dos Estados Unidos e de outros países, como parte do processo de apresentação de propostas de financiamentos de pesquisa.

Outra feita é no estágio planejar que a ferramenta é instalada e customizada, além de se mapearem processos e recursos para a execução de todo o plano de gestão em todas as fases do projeto. Pensar no plano de gestão é idealizar todo o processo e “começar com os objetivos do projeto (saídas desejadas, resultados e impactos) e trabalhar previamente, construindo uma pesquisa aberta, apoiada nas políticas de dados e planos de sustentabilidade” (WIGGINS et al., 2013, p. 3, tradução nossa).

Além da ferramenta de gestão, é recomendável o uso de um identificador persistente para cada pesquisador ou colaborador envolvido no projeto. Essa recomendação ajuda a garantir que as atribuições de autoria sejam agregadas aos dados de pesquisa no futuro. Durante a implantação do DMPTool, recomenda-se que o pesquisador (individual) proceda à geração do identificador *Open Researcher and Contributor Identification* (ORCID), visando à conexão dos dados de sua pesquisa. O identificador ORCID é composto por 16 (dezesseis) dígitos vinculados ao identificador (ID) do pesquisador e ao e-mail. Oferece um ID ORCID completo e o *link* para o seu registro público cingido à pesquisa e ao pesquisador.

Este estágio propicia o encaminhamento para a elaboração do plano de dados que encaminhará aos estágios subsequentes.

## SEGUNDO ESTÁGIO DO CICLO: COLETAR (*COLLECT*)

O estágio coletar refere-se à coleta primária dos dados desde a sua origem. Apoiar a infraestrutura tecnológica capaz de suportar diferentes tipos de dados (sensíveis, sigilosos, estáticos ou dinâmicos), bem como os fatores intrínsecos e

extrínsecos relacionados à natureza e às condições de coleta. As condições de coleta podem ser física, temporal e de acondicionamento. As coletas relacionadas com o tempo e o acondicionamento (preliminar) requerem atenção especial do pesquisador e dos colaboradores envolvidos.

Recomenda-se ter uma infraestrutura inicial com ferramentas de comunicação geral, a exemplo do *Dropbox*, *Google drive*, software específico ou equipamentos laboratoriais que acondicionam e asseguram a proteção dos dados primários. O pesquisador deve observar os aspectos relacionados com o volume, a extensão, a mídia dos dados, como também o *hardware*, software e outros recursos humanos envolvidos no processo. O volume estimado de armazenamento para os dados pode variar de *megabyte* (MB), *gigabyte* (GB), *terabyte* (TB) e *pentabyte* (PB); contudo, orienta-se estimar a previsão neste estágio em que os dados estão sendo coletados e armazenados preliminarmente.

Para a linguagem, recomenda-se a adoção de formatos legíveis por máquina, tais como *HyperText Markup Language* (HTML), *Extensible Markup Language* (XML), *Resource Description Framework* (RDF) e *Portable Document Format* (PDF-A), que também são úteis para manipular textos e documentos não estruturados, baseados em padrões abertos. O formato HTML é uma linguagem de marcação que auxilia na vinculação dos dados, conjuntos de dados (diferentes formatos, mídias e extensões) e documentos utilizando *hyperlinks* e *links* como âncoras para a vinculação das publicações, que podem ser dados, artigos, documentos ou outros suportes digitais. O formato XML é outra linguagem utilizada para a criação de documentos com dados organizados de forma hierárquica, que geram linguagens de marcação para necessidades especiais com a finalidade de compartilhar informações pela Internet.

Os arquivos *Comma-separated Values* (CSV) é outro padrão aberto aplicado para o formato textual que faz uma ordenação de *bytes* ou um formato de terminador de linha. Isso permite que os dados sejam manipulados por diferentes ferramentas, perdendo o contexto de limitação temporal pela obsolescência de software de acesso aos dados. Ele comumente é usado em software oficial, tais como *Microsoft Excel* e *Libre Office Calc*.

O padrão *Moving Picture Experts Group* (MPEG-4) é utilizado primeiramente para compressão de dados digitais de áudio e vídeo. O acrônimo *Tagged Image File Format* (TIFF) é um formato de arquivo para imagens digitais de arquivos gráficos com elevada definição de cores. E o *Joint Photographic Experts Group* (JPEG) é um padrão aberto para compressão de imagens de alta definição, facilitando a aplicação de diferentes ferramentas computacionais para a manipulação de imagens.

Aspectos relevantes são observados quanto aos procedimentos especiais, tais como: licenças, permissões, equipamentos, padrões, protocolos e repositórios que serão atendidos nos próximos estágios.



## TERCEIRO ESTÁGIO DO CICLO: ASSEGURAR (*ASSURE*)

Realizados os procedimentos preliminares de qualidade, validação e acondicionamento, ocorre a migração dos dados para uma ferramenta compatível que favoreça o compartilhamento, a curadoria e a preservação em interface única.

A adoção de identificadores vinculados aos dados, a exemplo do *Digital Object Identifier* (DOI), é uma prática recomendada nesse estágio. O DOI é um identificador persistente que agrega os dados (digitais), objetos (digitais) e documentos relacionados em uma mesma pesquisa. É um sistema que, além de assegurar a qualidade e a identificação, proporciona o gerenciamento dos direitos de propriedade intelectual, atribuições e citações de dados ou objetos digitais. Estabelece, ainda, a ordem de atribuição realizada para as citações de pesquisador(es), colaborador(es), instituições, agências de fomento e parceiros externos.

No que tange aos aspectos autorais, é relevante definir os procedimentos legais e quais licenças públicas serão adotadas, antes de avançar para o estágio seguinte. A definição de procedimentos legais, éticos, normativos e critérios para as condições e restrições de uso e *status* dos dados (aberto, fechado ou embargado) é recomendável e prevista em políticas e diretrizes desde o plano inicial. Os procedimentos devem promover a autoria de todos os envolvidos na pesquisa.

Além dos procedimentos já descritos, a infraestrutura de um repositório que sustente os estágios futuros de compartilhamento, curadoria e preservação é preliminarmente prevista no estágio assegurar. Recomenda-se também que a implantação da plataforma seja de código aberto e possibilite a agregação, as aplicações, a configuração do padrão de metadados, o protocolo de interoperabilidade e as ferramentas *web*.

## QUARTO ESTÁGIO DO CICLO: DESCREVER (*DESCRIBE*)

O estágio descrever é de extrema relevância para a descrição dos metadados de dados de pesquisa. A qualidade na descrição dos metadados e a adoção de padrões irão contribuir para a recuperação, o compartilhamento e a curadoria permanente. Os metadados são dados definidos como elementos que representam os dados do processo de investigação; sendo o padrão responsável pela descrição sistemática e acurada dos elementos.

Inúmeros padrões de metadados foram desenvolvidos para atender as especificidades das diversas disciplinas da área do conhecimento, com a finalidade de descrever minuciosamente e de forma compatível com a natureza dos dados (recursos). A *National Information Standards Organization* (NISO) define que “metadados são informações estruturadas que descrevem, explicam, localizam,

ou permitem que um recurso informacional possa ser fácil de recuperar, usar ou gerenciar” (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004, tradução nossa).

Os recursos no contexto de um padrão de metadados possuem conteúdo significativo (elementos) e diversos formatos (dados, informações ou objetos). É o padrão que auxilia na definição dos detalhes e a sistemática de descrição dos elementos de dados ou conjunto de dados contidos em um recurso informacional.

Na *e-Science*, as “informações sobre dados (metadados) são registradas para que outros possam descobrir, adquirir, interpretar e utilizar os dados de forma eficaz” (WIGGINS et al., 2013, p. 7, tradução nossa). A estruturação de metadados e a complexidade existente em diferentes disciplinas, por vezes, requerem o auxílio de recursos, técnicas e vocabulários controlados que apoiem o processo de descrição.

Por fim, esse é um estágio de grande relevância, pois assegura o compartilhamento, a preservação e o reuso dos dados em longo prazo; a continuidade do projeto nos estágios seguintes; e oportuniza que os metadados sejam processados em um repositório especializado.

Ademais, é nesse estágio que os procedimentos normativos e legais que promovem a recompensa autoral de pesquisadores, colaboradores e instituições envolvidas sejam assegurados.

## QUINTO ESTÁGIO DO CICLO: PRESERVAR (*PRESERVE*)

É nesse estágio que o plano de preservação para os dados primários em longo prazo é desenvolvido. A preservação tem a finalidade de proporcionar *backup* e guarda segura dos arquivos digitais. Utiliza recursos (manuais, eletrônicos e duplicados), ferramentas de armazenagem em nuvens, bem como o acondicionamento em software de gerenciamento eletrônico, laboratórios, estufas, *data centers* e repositórios especializados. Quando a preservação é realizada em repositório especializado, permite a curadoria em longo prazo e viabiliza os processos de conversão, reformatação e salvamento dos dados para o uso, reuso e reprodutibilidade futuros.

Esse estágio conduz a dois procedimentos importantes no gerenciamento dos dados:

- a) primeiro: está direcionado ao término da coleta dos dados primários, pois a identificação das tecnologias que suportam o *backup* em curto prazo assegura a preservação e guarda preliminar realizadas;
- b) segundo: direciona à armazenagem em longo prazo.

Contudo, salienta-se que a complexidade entre os diferentes tipos de dados requer padrões e protocolos que suportem a agregação, a curadoria e a interoperabilidade para estruturas complexas e abertas, com vistas à integração e à descoberta eficaz. É importante também adotar requerimentos específicos à proveniência e à tipologia dos dados, bem como à instalação aberta de licença pública, patente *on-line* e software. As licenças públicas abertas preveem a revisão e aprovação de licenciamento público para diferentes atribuições autorais (documentos, códigos, dados), bem como a previsão para não atribuições.

Na vertente do acesso aberto, a Rede Brasileira de Serviços de Preservação Digital - Cariniana, promovida pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), é um excelente exemplo de curadoria e preservação digital. A rede é responsável pela preservação digital compartilhada da informação científica, tecnológica e cultural e “busca disponibilizar serviços de preservação digital para instituições com publicações de acesso livre, além de mecanismos que facilitem a automatização dos processos de identificação, armazenamento, validação e conversão para novos formatos digitais” (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2015). Em 2015, a rede ampliou seu escopo para a preservação de dados de pesquisa com a integração ao Repositório *Dataverse*.

A rede disponibiliza políticas com orientações da *Harvard Dataverse* para as condições de acesso e permissões de uso. A política é orientada para aumentar a visibilidade dos dados da pesquisa por meio do reconhecimento e crédito acadêmico através da citação. As permissões e restrições adotadas para o compartilhamento e disponibilidade dos conteúdos e dados de pesquisa no repositório *Murray Archive* são gerenciadas por licenças no *DataCite*.

Em suma, observa-se que esse estágio não ocorre de forma simples e rápida, muito embora algumas iniciativas já disponibilizem soluções tecnológicas estruturadas e abertas para a descrição de metadados e protocolos de interoperabilidade dentro de repositórios, com base em padrões, tais como: OAI-PMH, XML, Dublin Core, DIDL ou DDI compatíveis com ambientes complexos e domínios específicos (WOUTERSEN-WINDHOUWER; BRANDSMA, 2009).

## SEXTO ESTÁGIO DO CICLO: DESCOBRIR (*DISCOVER*)

Após a garantia da curadoria e preservação em longo prazo, o estágio descobrir proporciona o completo compartilhamento, integração e visibilidade dos dados. A partir da integração com repositórios especializados, é também possível acessar diretórios de dados, *datasets*, buscadores especializados ou metabusca-

dores para a descoberta de dados, tornando os metadados localizáveis, detectáveis, reutilizáveis e citáveis.

No contexto internacional, existe uma gama de repositórios especializados em dados de pesquisa. Esses repositórios são oriundos de instituições de pesquisa e ensino, instituições governamentais e organizações não-governamentais. A seguir, alguns exemplos de repositórios especializados para o ambiente dos dados.

O *CiteBank* é um exemplo de repositório especializado de acesso aberto que faz a agregação de citações para artigos e publicações relacionados à biodiversidade. Ele fornece recursos de pesquisa e navegação para publicações armazenadas em vários repositórios internacionais e possibilita que pesquisadores compartilhem bibliografias especializadas.

O *DataONE Search* é uma plataforma que agrega repositórios. Os participantes agregadores são denominados de nós membros. *DataONE Search* é uma implementação personalizada do aplicativo de pesquisa de código aberto MetacatUI (<https://github.com/NCEAS/metacatui>), desenvolvido no Centro Nacional de Análise e Síntese Ecológica.

A *Global Change Master Directory* (GCMD) é um diretório mundial que contém metadados de coleção de dados e dados de pesquisa armazenados em repositórios especializados nas áreas das Ciências da Terra e Ambiental. Oferece pesquisa facetada e pesquisa de conjuntos de dados por extensão espaço-temporal, resolução, centro de dados, localização, instrumento, plataforma e projeto.

Alguns dos exemplos acima são apenas ilustrativos para a discussão em pauta, pois em projetos de dados se estimulam parcerias institucionais para maior visibilidade da pesquisa aberta e colaborativa mediante o compartilhamento de dados modelados a partir do propósito da pesquisa.

## SÉTIMO ESTÁGIO DO CICLO: INTEGRAR (*INTEGRATE*)

Muito embora a integração dos dados às vezes seja realizada preliminarmente em estágios anteriores, o estágio integrar potencializa a vinculação dos repositórios e a integração dos metadados, por meio dos protocolos de interoperabilidade, com a finalidade de prover maior visibilidade dos dados e da pesquisa, bem como novas análises e investigações para o reuso e a reprodutibilidade no futuro.

Nesse sentido, a agregação através dos protocolos de interoperabilidade possui dois propósitos:

- a) agregação interna dos diferentes dados ou conjunto de dados gerados na pesquisa científica em uma mesma interface;

- b) integração do projeto de dados com outras fontes externas, visando incentivar a publicação e compartilhamento da publicação tradicional avaliada por pares com a distribuição de dados, através de sites, catálogos de dados, mídia social e outros locais (FAUNDEEN et al., 2013, p. 3, tradução nossa).

Assim, além da integração aos diferentes canais de disseminação, também é possível proporcionar a visibilidade e a acessibilidade para diferentes usuários e contextos.

Inúmeros protocolos são adotados ao contexto de dados e objetos digitais, a exemplo do *Open Archives Initiative* (OAI). O OAI é o protocolo de interoperabilidade adotado para tornar eficiente a disseminação da pesquisa. E o *Open Archives Initiative Object Reuse and Exchange* (OAI-ORE)<sup>1</sup> é responsável pela definição de padrões, recursos e especificações que designam e auxiliam na troca de agregações de recursos da *web*. Tais agregações são adotadas para combinar os objetos distribuídos em vários tipos de mídias, recursos informacionais, dados e objetos digitais. O *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) deve ser adotado como uma opção técnica simples para prover a possibilidade de metadados interoperáveis e disponíveis para serviços baseados em padrões *web* abertos, tais como, o HTTP e o XML.

Em resumo, a agregação e o compartilhamento de dados ou de conjunto de dados se configura em uma etapa complexa no processo de gerenciamento, muito embora haja a tendência para reduzir custos, distância entre pesquisadores e instituições, bem como aumentar a visibilidade e garantir a continuidade de pesquisas importantes e complexas para o reuso e a reprodutibilidade futuras.

## OITAVO ESTÁGIO DO CICLO: ANALISAR (ANALYZE)

No modelo adotado, esse estágio é o último do ciclo de vida e destina-se à análise e às avaliações dos dados de pesquisa já compartilhados. As análises são realizadas por ferramentas analíticas, métricas especializadas, como também por software específico ou câmaras de compensação de dados.

O processo de pesquisa nesse estágio está completo, e o pesquisador ou grupo de pesquisa podem fazer derivações dos seus dados com a finalidade de atingir duas performances:

- a) utilizar os dados para análises que atendam os objetivos do projeto. Muitas ferramentas de software estão disponíveis para apoiar as atividades de exploração, análise e visualização dos dados. A *análise dos dados* é feita por computadores, *grid* ou *cloud computing*, que promovem tarefas

1 Disponível em: <<http://www.openarchives.org/ore/>>. Acesso em: 7 mar. 2015.

estatísticas, execução de modelos, estimativas de parâmetros, gráficos e plotagens;

- b) proporcionar a completa reprodutibilidade dos dados que é a chave para os métodos científicos e processos complexos reproduzidos. Recomenda-se que haja a completa e eficiente preservação de uma boa documentação como pré-requisito essencial para a reprodutibilidade eficaz.

Ferramentas que documentam o *workflow* e facilitam a reprodutibilidade são adotadas, a exemplo do *VisTrails* e do projeto *Kepler*. O *VisTrails* é um sistema de código aberto para o gerenciamento de fluxo de trabalho e processos de conteúdos científicos que oferece suporte à exploração e à visualização de dados, desenvolvido no *Scientific Computing and Imaging Institute* da Universidade de Utah. Além de fornecer uma infraestrutura geral que pode ser combinada com repositórios, sistemas e bibliotecas, também suporta a exploração e a visualização de dados com versões pré-compiladas para Windows, Mac OS X e Linux.

O projeto *Kepler* promove e apoia as capacidades, o uso e a conscientização da aplicação de fluxos de trabalho científicos. O *Kepler* é de código aberto e livre, desenvolvido e mantido pela colaboração entre equipes compostas pela *University of California de Davis, Santa Barbara e San Diego* e pela *NSF - Funded Kepler/CORE team*. Foi projetado para auxiliar cientistas, analistas e programadores de computadores, criar, executar e compartilhar modelos e análises em uma ampla gama de disciplinas científicas.

A aplicação de ferramentas de alta performance para análise métrica e visualização de dados é uma prática recomendada. O estágio também apoia iniciativas e abordagens que descrevem e associam todo o conteúdo de uma pesquisa em um mecanismo de leitura óptica para que ele possa ser mais facilmente compartilhado e permutado para além do reuso, oportunizando a reprodutibilidade de dados e códigos abertos. Analisar e visualizar os dados são processos de qualidade e validação da ciência aberta.

A Figura 2 é uma representação de soluções tecnológicas acessíveis, abertas e recomendáveis para cada estágio do ciclo.

Em suma, o ciclo de vida dos dados é uma metodologia que sistematiza todo o processo de gestão de dados em uma ambiência aberta e compartilhada, além de proporcionar a ciberinfraestrutura tecnológica adequada a cada estágio. Muito embora cada instituição ou grupo de pesquisa opte por adotar ou desenvolver um modelo de ciclo de vida adequado aos propósitos da pesquisa e à natureza dos dados, é a ciberinfraestrutura que dá sustentabilidade aos estágios durante todo o processo de gestão.



Figura 2 – Ciberinfraestrutura baseada no ciclo de vida dos dados  
Fonte: Autoria própria (2016).

Afinal, o princípio básico da ciência aberta é tornar os dados de pesquisa, bem como documentos, ferramentas e aplicações tecnológicas geradas no processo de pesquisa, acessíveis, interoperáveis, abertos, reutilizáveis e reproduzíveis. Políticas e diretrizes devem conduzir o projeto de dados em consonância com o propósito da pesquisa, instituições envolvidas e em conformidade com padrões internacionais.

## PADRÕES DE METADADOS PARA REPOSITÓRIOS DE DADOS

A seção anterior apresentou a ciberinfraestrutura que apoia o ciclo de vida dos dados. Esta seção traz como destaque o estágio descrever, relacionado à descrição de metadados com base em padrões.

O padrão de metadados no processo de pesquisa é usado para realizar a representação adequada e padronizada de recursos armazenados para interoperabilidade nos estágios subsequentes. Teoricamente, os padrões de metadados são divididos em três níveis (Quadro 1).

Esses níveis realizam a descrição do recurso de acordo com sua complexidade e especificidade. A natureza do recurso e o grau de detalhamento irão determinar o nível do padrão a ser adotado. Importante salientar que, quanto maior for o grau de detalhamento da descrição, maior será a recuperação semântica dos dados de pesquisa.

Nível 1	São metadados com padrões simples e consiste de dados desestruturados que são, em alguns casos, extraídos automaticamente por robôs automatizados. Na maioria dos casos possuem uma redução semântica.
Nível 2	São padrões de metadados estruturados que constituem e definem os metadados em um esquema. Eles são baseados em padrões que fornecem uma descrição mínima do recurso para a identificação, localização e valorização.
Nível 3	São os padrões de metadados ricos, considerados altamente estruturados. Eles são constituídos por metadados complexos e apresentam uma descrição mais detalhada formal e de estrutura. Eles são baseados em normas e códigos de um domínio específico especializado, permitindo a descrição de um recurso de informação individual ou pertencente a uma coleção. Facilitando a localização, recuperação e interoperabilidade de recursos de informação.

Quadro 1 – Níveis dos padrões de metadados

Fonte: Alves et al. (2014, p. 155 apud OLIVEIRA 2016).

Inicialmente os padrões foram desenvolvidos para descrever os recursos impressos contidos em bibliotecas, museus, arquivos e centros de documentação. Com a evolução tecnológica, padrões foram sendo adaptados para a descrição de recursos digitais em formato eletrônico e aplicáveis em sistemas automatizados, bases de dados e gerenciadores de periódicos eletrônicos. Atualmente, os padrões evoluíram para se adequarem à descrição de dados de pesquisa ou conjuntos de dados (*datasets*) e aos repositórios especializados e digitais.

Para a descrição dos metadados e a escolha de um padrão compatível com a natureza dos dados, adotam-se critérios para assegurar a qualidade da descrição e o preenchimento completo dos mesmos. Dentre eles, recomenda-se elaborar um *check list* que responda às perguntas: *why, who, what, when, where* e *how* (Quadro 2) para a descrição eficiente dos metadados.

Seguir procedimentos de qualidade, descrição, confidencialidade, fidedignidade e escolher um padrão de metadados específico, são recomendações internacionais que assegurarão a conformidade internacional.



PERGUNTAS	ORIENTAÇÕES
<i>Why</i> (Por quê?)	Refere-se ao entendimento dos dados e o contexto em que eles foram criados. Os detalhes devem incluir as questões científicas ou proposições para a coleta de dados, razões que envolvem contribuidores voluntários e uma visão geral sobre o conjunto de dados.
<i>Who</i> (Quem?)	Trata-se da descrição dos metadados relacionados com quem coletou os dados ou quem patrocinou a coleta. Ele também deve especificar quem deverá ser contactado com perguntas sobre os dados e como deve ser citado o conjunto de dados para garantir que todos os envolvidos obtenham reconhecimento na pesquisa. Este é um espaço que deve descrever os metadados de todos os pesquisadores, colaboradores, agências e instituições que se relacionaram à pesquisa.
<i>What</i> (O quê?)	Requer uma descrição apurada dos dados. Foram desenvolvidas algumas categorias que ajudam a nortear a descrição. Com base no <i>Data Management Guide for Public Participation in Scientific Research</i> , as categorias recomendam seis passos para a descrição desta questão, relacionadas ao <b>contexto digital</b> (nomes de arquivos, formatos de arquivos, modificações de datas, arquivos de exemplo, conjuntos de dados relacionados e procedimentos de processamento de dados); <b>Parâmetro/ detalhes variáveis</b> (unidades de medida, formatos de dados, precisão e exatidão); <b>Informações sobre dados</b> (taxonomias, codificação, procedimentos); <b>Conteúdo de arquivos de dados</b> (definições de parâmetros e explicação de seus formatos, avaliação de qualidade notas); <b>Informações taxonômicas adicionais</b> (usando taxonomias padrão quando possível); <b>Organização dos dados</b> (relacionamentos entre entidades de dados, arquivos, diretórios de banco de dados).
<i>When</i> (Quando?)	Relaciona-se ao detalhamento dos aspectos temporais, como período (ano, mês, dia), tempo, regularidade e intervalos da coleta de campo.
<i>Where</i> (Onde?)	Direciona-se ao detalhamento dos aspectos de cobertura geográfica e espacial relacionados com o país, a cidade, local, coordenadas espaciais (latitude e longitude) e condições meteorológicas (clima, temperatura).
<i>How</i> (Como?)	Estabelece as condições de criação dos dados. Os detalhes são muito importantes para as possibilidades de reuso e interpretação futura dos dados. O detalhamento de protocolos, técnicas, métodos, diagramas, software, modelos, esquemas e estatísticas utilizadas nos procedimentos metodológicos e de coleta de campo que originou os dados é relevante para uma descrição qualitativa dos metadados. Documentos adicionais aos dados também devem ser informados e se possível vinculados ao dado ou conjunto de dados.

Quadro 2 – Orientações para a descrição de metadados

Fonte: Autoria própria (2016).

Para complementar este tópico, a seção seguinte apresenta um levantamento de padrões com ênfase em domínios específicos - ecologia, biodiversidade, física, genoma - da ciência, desenvolvidos para práticas com dados de pesquisa.

## MODELOS DE PADRÕES DE METADADOS COM ÊNFASE NO DOMÍNIO ESPECÍFICO

O Quadro 3 apresenta um levantamento de padrões de metadados com ênfase em domínios específicos e aplicados para a descrição de dados de acordo com a natureza da pesquisa e área do conhecimento. A importância de identificar padrões específicos a domínios especializados, em especial nas disciplinas complexas e de grande volume de dados gerados nos processos de investigação, foi uma das mudanças ocorridas a partir de práticas na *e-Science*.

Importante destacar que vários padrões são desenvolvidos por instituições de pesquisa, governamentais e não-governamentais, para atender necessidades específicas. À guisa de ilustração, veja-se o Quadro 3.

O padrão *Content Standard for Digital Geospatial Metadata* (CSDGM), estruturado de acordo com a ISO, possui ênfase em dados geoespaciais e foi desenvolvido pelo *Federal Geographic Data Committee* (FGDC) pertencente à organização governamental americana *U.S. Geological Survey* (USGS).

O padrão *Ecological Metadata Language* (EML) tem ênfase nos dados ecológicos e foi desenvolvido pela rede de conhecimento *Knowledge Network for Bio-complexity* (KNB), que dá sustentabilidade para a biocomplexidade da pesquisa ambiental e ecológica por meio de um repositório aberto.

O padrão *Darwin Core Biodiversity Information Standards* (TWDG) possui sua ênfase em espécimes naturais. É destinado à descrição de metadados de objetos contidos em coleções de espécimes da história natural e em bancos de dados de observação das espécies, desenvolvido pelo *Natural History Museum and Biodiversity Research Center*.

O *Genome Metadata* é um padrão com ênfase em dados científicos de genomas individuais. O *Pathosystems Resource Integration Center* (PATRIC) consiste em 61 (sessenta e um) diferentes campos de metadados, chamados atributos, organizados em 07 (sete) grandes categorias para descrição.

O *Isa Tools* é a suíte do *ISA-Tab Metadata* que tem ênfase na coleta e comunicação de metadados complexos desenvolvidos pela Nature. É um padrão de alta complexidade e atende a descrição de metadados de dados científicos e de publicações desenvolvidas pela Nature. A Nature é uma plataforma que hospeda aproximadamente 80 revistas e agrega milhões de cientistas do mundo todo, com aces-

so aberto, publicação *on-line* contínua e uma experiência de agregação de dados científicos com o artigo principal, fornecendo possibilidades de acesso ao conjunto dos dados de pesquisas. O acesso a esses dados ocorre por meio de descritores de dados, a partir da combinação da publicação científica e informações estruturadas em *datasets*, para maximizar e reutilizar os dados de pesquisa.

O padrão *Statistical Data and Metadata Exchange* (SDMX) é uma iniciativa para promover normas para o intercâmbio, agregação de dados e ferramentas de repositórios de informações estatísticas oficiais.

ORGANIZAÇÃO	PADRÕES DE METADADOS	ÊNFASES
U.S. GEOLOGICAL SURVEY (USGS)	<i>Content Standard for Digital Geospatial Metadata (CSDGM)</i>	Ênfase em dados geoespaciais.
	<b>Website:</b> < <a href="http://www.fgdc.gov/metadata/geospatial-metadata-standards">http://www.fgdc.gov/metadata/geospatial-metadata-standards</a> >	
	<i>Geographic Information: Metadata ISO 19115/19139.</i>	Ênfase em dados e serviços geoespaciais.
	<b>Website:</b> < <a href="http://www.fgdc.gov/metadata/geospatial-metadata-standards#fgdcendorsedisostandards">http://www.fgdc.gov/metadata/geospatial-metadata-standards#fgdcendorsedisostandards</a> >	
	<i>Geography Markup Language (GML)</i>	Ênfase em dados geográficos (rotas, rodovias e pontes) em linguagem XML.
	<b>Website:</b> < <a href="http://www.opengeospatial.org/standards/gml">http://www.opengeospatial.org/standards/gml</a> >	
	<i>The Federal Geographic Data Committee (FGDC)</i>	Ênfase em padrão de metadados para conteúdos digitais geoespaciais.
KNOWLEDGE NETWORK FOR BIOCOMPLEXITY (KNB)	<b>Website:</b> < <a href="http://www.fgdc.gov/">http://www.fgdc.gov/</a> >	
	<i>Ecological Metadata Language (EML)</i>	Ênfase em dados ecológicos.
	<b>Website:</b> < <a href="http://www.dcc.ac.uk/resources/metadata-standards/eml-ecological-metadata-language">http://www.dcc.ac.uk/resources/metadata-standards/eml-ecological-metadata-language</a> >	

ORGANIZAÇÃO	PADRÕES DE METADADOS	ÊNFASES
NATURAL HISTORY MUSEUM AND BIODIVERSITY RESEARCH CENTER	<i>Darwin Core - Biodiversity Information Standards (TWDG)</i>	Ênfase em espécies de museus. Um padrão para descrição de objetos que contêm coleções de espécimes de história natural e conjunto de dados de observação de espécies.
	<b>Website:</b> < <a href="http://rs.tdwg.org/dwc/">http://rs.tdwg.org/dwc/</a> >	
DataCite CONSORTIUM	<i>DataCite Metadata Schema</i>	Esquema de padrão de metadados com propriedades para a segurança e identificação consistente de dados para citação e recuperação.
	<b>Website:</b> < <a href="http://www.dcc.ac.uk/resources/metadata-standards/datacite-metadata-schema">http://www.dcc.ac.uk/resources/metadata-standards/datacite-metadata-schema</a> >	
DDI ALLIANCE	<i>Data Documentation Initiative (DDI)</i>	Padrão internacional de fácil uso para descrever dados comportamentais, sociais e ciências econômicas.
	<b>Website:</b> < <a href="http://www.dcc.ac.uk/resources/metadata-standards/list?page=1">http://www.dcc.ac.uk/resources/metadata-standards/list?page=1</a> >	
GENOME METADATA OVERVIEW	<i>Genome Metadata</i>	Ênfase em metadados que são dados descritivos sobre genomas individuais. Metadados genoma em PATRIC consistem em 61 diferentes campos de metadados, chamados atributos, organizados em sete grandes categorias.
	<b>Website:</b> < <a href="http://enews.patricbrc.org/faqs/2-genome-data-and-tools/genome-metadata-faqs/">http://enews.patricbrc.org/faqs/2-genome-data-and-tools/genome-metadata-faqs/</a> >	
UNIVERSITY OF OXFORD ISA-TAB	<i>ISA-Tab</i>	Ênfase na coleção e comunicação de metadados complexos para experimentos "omics-based" e a combinação de tecnologias.
	<b>Website:</b> < <a href="http://isatab.sourceforge.net">http://isatab.sourceforge.net</a> >	

ORGANIZAÇÃO	PADRÕES DE METADADOS	ÊNFASES
MIBBI INITIATIVE	Minimum Information for Biological and Biomedical Investigations (MIBBI)	Ênfase em informações mínimas para várias disciplinas biológicas.
	<b>Website:</b> < <a href="http://www.dcc.ac.uk/resources/metadata-standards/list?page=2">http://www.dcc.ac.uk/resources/metadata-standards/list?page=2</a> >	
CALIFORNIA DIGITAL LIBRARY (CDL)	CopyrightMD	Ênfase em um esquema de XML para metadados de direitos autorais.
	<b>Website:</b> < <a href="http://www.cdlib.org/groups/rmg/">http://www.cdlib.org/groups/rmg/</a> >	
ANSI/NISO	<i>Data Dictionary - Technical Metadata for Digital Still Images (Z39.87-2006)</i>	Ênfase em elementos de dados técnicos relevantes ao gerenciamento de imagens fixas digitais.
	<b>Website:</b> < <a href="http://www.niso.org/kst/reports/standards?step=2&amp;gid=None&amp;project_key=b897b0cf3e2ee526252d9f830207b3cc9f3b6c2c">http://www.niso.org/kst/reports/standards?step=2&amp;gid=None&amp;project_key=b897b0cf3e2ee526252d9f830207b3cc9f3b6c2c</a> >	
LIBRARY OF CONGRESS	<i>Metadata for Images in XML Standard (MIX)</i>	Ênfase em metadados para imagens fixas em esquema XML.
	<b>Website:</b> < <a href="http://www.loc.gov/standards/mix/">http://www.loc.gov/standards/mix/</a> >	
DUBLIN CORE INITIATIVE	<i>Dublin Core Metadata</i>	Ênfase em publicações e recursos web.
	<b>Website:</b> < <a href="http://dublincore.org/documents/dces/">http://dublincore.org/documents/dces/</a> >	

Quadro 3 – Padrões de metadados com ênfase em domínios específicos

Fonte: Autoria própria (2016).

Adiciona-se, por fim, o padrão *Preservation Metadata Maintenance Activity* (PREMIS) que funciona como um dicionário de dados que apoia a preservação de metadados de objetos digitais e assegura a sua usabilidade a longo prazo.

Quando a descrição dos metadados não atende aos procedimentos de qualidade e conformidade internacional pré-definidos no plano de pesquisa, os metadados relacionados não são interoperados em repositórios de dados. Daí a relevância de atender critérios, diretrizes e boas práticas disponíveis por instituições envolvidas com a pesquisa aberta. O uso de vocabulários controlados, tesouros e ferramentas automáticas de geração de *codes* de metadados também é recomendável. Ademais, questões autorais relacionadas com as atribuições de autoria dos envolvidos – pesquisadores, grupos de pesquisa, instituições e agências de fomento – são adotadas no estágio descrever.

Por fim, adotar procedimentos e padrão de metadados destinados ao domínio específico de uma pesquisa compartilhada e aberta no contexto da ciência aberta é de suma importância para garantir que a pesquisa se perpetue para as futuras décadas.

## CONSIDERAÇÕES FINAIS

Como visto, a pesquisa em um contexto de ciência aberta assume novos contornos e difere das práticas nos moldes tradicionais. A *práxis* é orientada por dados de pesquisa que são os produtos primários da investigação.

No estudo em questão, optou-se por conceituar ciência aberta como um termo amplo que contempla políticas, diretrizes e procedimentos que conduzem à *práxis* da *e-Science*, ao passo que a *práxis* da *e-Science* apresenta um novo fazer e publicar da pesquisa científica, com características singulares (aberta, compartilhada e colaborativa) apoiada em três eixos: modelo de ciclo de vida dos dados, que garantem o completo processo de gestão; a ciberinfraestrutura, que dá sustentabilidade às soluções tecnológicas e técnicas para uso, reuso e reprodutibilidade dos dados primários em longo prazo; e os aspectos voltados ao compartilhamento e à colaboração científica que acomoda diferentes instituições, diversos pesquisadores e colaboradores e propósitos de pesquisa distintos.

Os aspectos tecnológicos envolvem o agrupamento de plataformas abertas, aplicações, padrões de metadados, protocolos de interoperabilidade, identificadores persistentes, linguagens compatíveis *web* e repositórios especializados para tornar os dados acessíveis, persistentes, confiáveis, detectáveis e reusáveis.

Os aspectos metodológicos atendem aos procedimentos científicos e suportam a colaboração científica e as diferentes formas de publicação - fontes de dados, informações, pesquisas, *links*, resultados, mídias e simulações - com os públicos (pesquisador, colaborador, produtor, editor e usuário), possibilitando a convergência, interatividade e o compartilhamento. Além disso apoiam o processo de pesquisa desde o estágio de planejamento, contemplando o rigor científico e procedimentos de domínios específicos.

Os aspectos técnicos relacionam-se a normas, padrões, vocabulários controlados e outros recursos que buscam a qualidade nos procedimentos e a conformidade internacional.

Por fim, o estudo buscou ir além dos aspectos relacionados aos padrões de metadados em repositórios digitais, pois o processo de uma pesquisa aberta colaborativa no escopo da ciência aberta transcende um único aspecto, constituindo-se em muitas dimensões. Outrossim, as práticas no escopo da ciência aberta visam também ao desenvolvimento econômico, social, cultural e o empoderamento dos envolvidos e da sociedade em geral.

## REFERÊNCIAS

- ALBAGLI, S. Ciência aberta em questão. In: ALBAGLI, S.; MACIEL, M. L.; ABDON, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015, p. 9-26. Disponível em: <<https://www.cienciaaberta.net/ciencia-aberta-questoes-abertas-o-livro/>>. Acesso em: 19 set. 2016.
- AMBINDER, D. M.; MARCONDES, C. H. Artigos científicos digitais na web: novas experiências para apresentação, acesso e leitura. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2012, Rio de Janeiro. **Anais eletrônicos...** Rio de Janeiro: ENANCIB, 2012. Disponível em: <<http://enancib.ibict.br/index.php/enancib/xii/enancib/paper/viewFile/3851/2974>>. Acesso em: 11 jul. 2016.
- ATKINS, D. E. et al. **Revolutionizing science and engineering through cyber infrastructure**: report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyber infrastructure. Washington, DC: [s. n.], 2003. Disponível em: <<https://www.nsf.gov/cise/sci/reports/atkins.pdf>>. Acesso em: 15 abr. 2015.
- BORGMAN, C. L. **From Gutenberg to the global information infrastructure**: access to information in the Networked World. London: MIT Press, 2003. Disponível em: <<https://mitpress.mit.edu/books/gutenberg-global-information-infrastructure>>. Acesso em: 19 abr. 2016.
- BORGMAN, C. L. Research data: who will share what, with whom, when, and why? In: CHINA-NORTH AMERICAN LIBRARY CONFERENCE, 5., 2010, Beijing. **Anais eletrônicos...** Disponível em: <<https://works.bepress.com/borgman/238/>>. Acesso em: 20 maio 2015.
- BORGMAN, C. L. **Scholarship in the digital age**: information, infrastructure, and the Internet. Cambridge: The MIT Press, 2007.
- COSTA, M. M.; CUNHA, M. B. da. A literatura internacional sobre e-science nas bases de dados LISA e LISTA. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 20, n. 44, p. 127-144, nov. 2015. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2015v20n44p127>>. Acesso em: 06 out. 2016.
- DIGITAL CURATION CENTRE. **DCC Curation Lifecycle Model**. Disponível em: <<http://www.dcc.ac.uk/resources/curation-lifecycle-model/>>. Acesso em: 13 set. 2016.
- FAUNDEEN, J. L. et al. **The United States geological survey science data lifecycle model**. Open-File Report, 2013. Disponível em: <<http://dx.doi.org/10.3133/ofr20131265>>. Acesso em: 13 set. 2016.
- FROSIO, G. F. **Open access publishing**: a literature review. Reino Unido: Research Fellow School of Law University of Nottingham, 2014. Disponível em: <<http://www.create.ac.uk/publications/open-access-publishing-a-literature-review/>>. Acesso em: 08 out. 2016.
- GRAY, J. eScience: a transformed scientific method. In: STEWART, T. et al. **The fourth paradigm**: data-intensive scientific discovery. Redmond, WA: Microsoft research, 2009. Disponível em: <<http://digital.library.unt.edu/ark:/67531/metadc31516/>>. Acesso em: 25 abr. 2016.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Rede Cariniana**. Disponível em: <<http://cariniana.ibict.br>>. Acesso em: 06 nov. 2015.

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH. **Guide to social science data preparation and archiving**: introduction. 2012. Disponível em: <<http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/index.html#cycle>>. Acesso em: 21 fev. 2015.

NATIONAL INFORMATION STANDARDS ORGANIZATION (NISO). **Standards. 2004**. Disponível em: <[http://www.niso.org/apps/group\\_public/projects.php](http://www.niso.org/apps/group_public/projects.php)>. Acesso em: 07 mar. 2016.

OLIVEIRA, A. C. S. de et al. Quarto paradigma científico: adoção do data life cycle model para o gerenciamento de dados científicos abertos do experimento e-Quilt Prototype. In: CONFERÊNCIA-LUSO BRASILEIRA DE ACESSO ABERTO, 6., 2015, Salvador. **Anais eletrônicos...** Disponível em: <<https://drive.google.com/folderview?id=0B2CiJWPT1NhwcNiZ1BwZWY5TVU&usp=sharing>>. Acesso em: 20 jul. 2016.

OLIVEIRA, A. C. S. de. **Desvendando a autoridade colaborativa na e-Science sob a ótica dos direitos de propriedade intelectual**. 2016. 300 f. Tese (Doutorado em Ciência da Informação) - Universidade Federal da Paraíba, João Pessoa, 2016. Disponível em: <<http://tede.biblioteca.ufpb.br:8080/handle/tede/8849>>. Acesso em: 01 maio 2017.

SALES, L. F.; SAYÃO, L. F.; SOUZA, R. F. Publicações ampliadas: um novo modelo de publicação acadêmica para o ambiente de e-science. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais eletrônicos...** Disponível em: <<http://enancib.ibict.br/index.php/enancib/xivenancib/paper/viewFile/4393/3516>>. Acesso em: 20 jan. 2016.

TENOPIR, C. et al. Data sharing by scientists: practices and perceptions. **Plos One**, v. 6, n. 6, p. 1-21, 29 jun. 2011. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>>. Acesso em: 25 ago. 2016.

TENOPIR, C.; KING, D.W. Communication patterns of engineers. [S. l.]: **Wiley-IEEE Computer Society Pr**, 2004. Disponível em: <<http://dl.acm.org/citation.cfm?id=966251>>. Acesso em: 12 set. 2015. ISBN:047148492X (Book).

WIGGINS, A. et al. **Data management guide for public participation in scientific research**. 2013. Disponível em: <<https://www.dataone.org/sites/all/documents/DataONE-PPSR-DataManagementGuide.pdf>>. Acesso em: 15 jun. 2015.

WOUTERSEN-WINDHOUWER, S.; BRANDSMA, R. Enhanced publications, state of the art. In: VERNOOY-GERRITSEN, M. (Ed.). **Enhanced publications**: linking publications and research data in digital repositories. Amsterdam: Amsterdam University Press, 2009. p. 19-91. Disponível em: <<http://dare.uva.nl/cgi/arno/show.cgi?fid=150723>>. Acesso em: 01 ago. 2017.



# DESAFIOS À COMUNIDADE IBERO-AMERICANA DE METADADOS EM REPOSITÓRIOS DIGITAIS PARA MAXIMIZAÇÃO DA INTEROPERABILIDADE

Ana Alice Baptista





## INTRODUÇÃO

Acesso Aberto (AA) significa “acesso em linha permanente, imediato e gratuito ao texto integral de artigos de investigação por qualquer pessoa em qualquer ponto da web (HARNAD, 2007). Os repositórios digitais científicos são uma das formas mais comuns de implementar o AA; constituem o que normalmente se designa por via verde para o AA. Estes repositórios normalmente incorporam vários tipos de documentos científicos, mas os mais comuns são os artigos científicos e as teses e dissertações. A informação sobre estes documentos é armazenada em registos que se expõem ao exterior como registos de metadados em formato OAI-DC, que mais não é do que um formato XML criado no âmbito da *Open Archives Initiative* (OAI) para veicular metadados utilizando os famosos 15 termos do *Dublin Core Metadata Element Set* (DCMES) (DUBLIN CORE METADATA INITIATIVE, 2008). Estes registos podem depois ser coletados por outro software que obedeça ao protocolo *Open Archives Initiative – Protocol for Metadata Harvesting* (OAI-PMH) de modo a serem incorporados em bases de dados que suportam serviços de valor acrescentado à comunidade científica.

A utilização do formato OAI-DC para os registos de metadados garante interoperabilidade sintática. Contudo, qualquer breve análise desses registos revela de forma límpida os problemas que ainda subsistem em termos de qualidade dos metadados e, por isso, da interoperabilidade semântica. Pretende-se com este documento de posição apresentar à comunidade Ibero-Americana de metadados em repositórios digitais as soluções já há vários anos divisadas pelas comunidades internacionais de metadados e de web semântica. Estas passam pela aplicação rigorosa do quarto nível de interoperabilidade *Dublin Core* (DC), o que implica o desenvolvimento e a aplicação, em *Resource Description Framework* (RDF), de perfis de aplicação de metadados específicos. Desafia-se, assim, a comunidade Ibero-Americana de metadados em repositórios digitais a dar um passo em frente rumo à web semântica e aos Dados Abertos Ligados (DAL)<sup>1</sup> a fim de maximizar a interoperabilidade semântica entre os seus repositórios e entre estes e os do resto do mundo.

Este capítulo está dividido em cinco seções. Após esta introdução, apresenta-se um breve contexto sobre o paradigma *Linked Data*, após o que se disserta brevemente sobre a qualidade dos metadados. Por fim, apresenta-se a proposta a efetuar à comunidade Ibero-Americana de repositórios digitais, apresentando a posição do artigo e terminando com as considerações finais.

1 Tradução livre da expressão em inglês *Linked Open Data* (LOD).

## LINKED DATA

*The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF) (WORLD WIDE WEB CONSORTIUM, 2013a).*

A web semântica, também chamada de web de dados (BERNERS-LEE, 2009b), contrapõe-se à web de documentos, no sentido em que é inteligível por máquinas, ao contrário da segunda, inteligível por humanos. Informação inteligível por máquinas é associada a informação inteligível por humanos, possibilitando às máquinas a associação, relacionamento e processamento dessa mesma informação a um nível global e interoperável. A web semântica não é outra web, nem sequer uma nova versão da web atual: trata-se, antes, do **acoplamento** à web atual de um nível de significado que é processável por máquinas. a web atual não deixa de existir – apenas é estendida e enriquecida com significado.

Trata-se simplesmente de ter dados estruturados seguindo rigorosamente normas *de jure* e *de facto* específicas a domínios ou que os atravessam. É o caso das normas relativas aos termos DC ou à forma como estes termos e os valores com eles relacionados são codificados, conforme recomenda o *World Wide Web Consortium* (W3C).

Tim Berners-Lee (2009a) denomina de dados abertos cinco estrelas os que são veiculados em RDF e estão ligados, isto é, têm contexto acoplado (Figura 1). Ou seja, não basta que os dados estejam codificados em RDF e tenham identificadores; é necessário que seja possível que as máquinas os **entendam** sem intervenção humana adicional, é necessário dar-lhes contexto para que sejam semanticamente interoperáveis e, por isso, processáveis.

Também preocupada com as questões da interoperabilidade semântica a nível global, a *Dublin Core Metadata Initiative* (DCMI) lançou em 2009 uma recomendação intitulada *Interoperability Levels for Dublin Core Metadata* (NILSSON; BAKER; JOHNSTON, 2009). Esta recomendação, já traduzida para o português, apresenta quatro níveis de interoperabilidade (Figura 2) e define como quarto nível, o mais alto, aquele em que exista um perfil de aplicação que empregue restrições aos registos (por exemplo, especifique como interpretar uma data ou quais os vocabulários controlados aceites para identificar o assunto ou o tipo do documento).

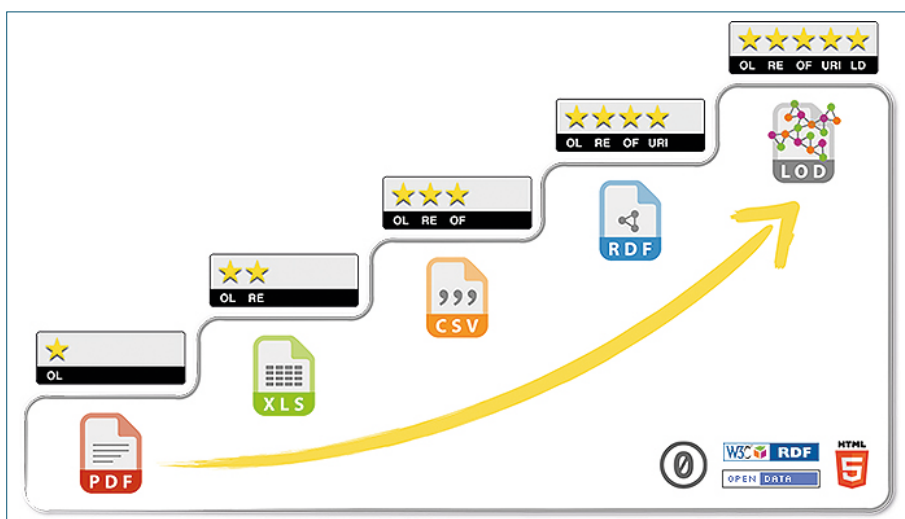


Figura 1 – Dados abertos cinco estrelas  
Fonte: Hausenblas (2012).

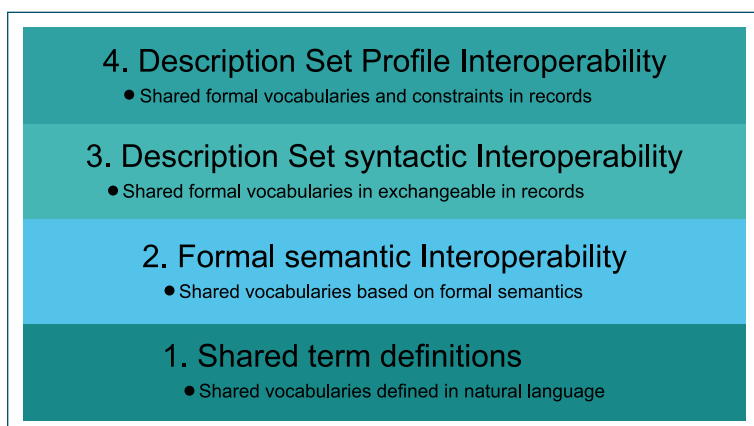


Figura 2 – Níveis de interoperabilidade *Dublin Core*  
Fonte: Nilsson, Baker e Johnston (2009).

A definição de perfil de aplicação de metadados não é consensual na literatura. O Enquadramento de Singapura define-o como “um pacote de documentação” composto pelos seguintes componentes: requisitos funcionais, modelo de domínio, perfil de conjuntos de descrições, guias de utilização e guias de codificação (NILSSON; BAKER; JOHNSTON, 2008). Contudo, a definição que parece ter tido mais aceitação na comunidade é a de Heery e Patel (2000), que definem perfil da aplicação de metadados como “esquemas que consistem em elementos de dados provenientes de um ou mais *namespaces*, combinados pelos implemen-

tadores e otimizados para uma aplicação local em particular”. Esta definição é relativamente antiga e não oferece aos perfis de aplicação de metadados o poder expressivo dos perfis de conjuntos de descrições do Enquadramento de Singapura, tal como definido por Coyle e Baker (2009) e por Nilsson (2008). No entanto, é bastante simples e clara, razão pela qual é adotada neste capítulo.

São raros os repositórios digitais que têm definidos perfis de aplicação ou que sequer usam esquemas de metadados para além do DC simples, o OAI-DC (ANDRADE; BAPTISTA, 2015). Mesmo utilizando apenas OAI-DC, os repositórios não explicitam claramente como interpretar os valores associados às propriedades em cada um dos seus registos OAI-DC. Além disso, não há coerência entre registos de repositórios diferentes. É até possível que em alguns repositórios nem sequer haja coerência através dos seus próprios registos (aqui está um bom tema de investigação para um mestrando). A seção seguinte apresenta alguns exemplos e respetiva discussão sobre este assunto.

## QUALIDADE DOS METADADOS

Uma consulta rápida aos registos OAI-DC dos repositórios digitais mostra claramente que há casos em que:

- a) a semântica das propriedades não está a ser respeitada. Por exemplo, existem casos em que os valores *Open Access* ou *Restrict Access* aparecem associados à propriedade *dc:rights*, quando esta se refere aos direitos legais sobre o recurso e não ao tipo de acesso. Outro exemplo é a propriedade *dc:identifier* ser usada não para fornecer um identificador do recurso e sim outra informação como, por exemplo, a área disciplinar a que o recurso diz respeito;
- b) não são utilizadas propriedades diferentes para veicular informação de natureza diferente. Um exemplo muito frequente é a repetição da propriedade *dc:date* com valores distintos, mas sem informação sobre o que cada uma das datas significa. Claramente, nestes casos será necessário utilizar propriedades diferentes. No caso particular do *dc:date*, pode-se utilizar refinamentos definidos pela própria DCMI, tais como *dct:created*, *dct:modified* ou *dct:issued*. O serviço *Linked Open Vocabularies* (LOV) permite a pesquisa de informação relativa a vários vocabulários abertos ligados, pelo que propriedades não existentes no DC Terms (DUBLIN CORE METADATA INITIATIVE, 2012), podem ser ali pesquisadas e identificadas;
- c) os valores associados às propriedades não estão preparados para serem facilmente interpretados fora do âmbito estrito do repositório a que dizem respeito. Por exemplo, muitos repositórios utilizam os idiomas

e alfabetos oficiais dos países onde estão sediados para representar o assunto (*dc:subject*), o tipo de documento (*dc:type*), os direitos (*dc:rights*), entre outros. Alguns repositórios até usam texto para representar os valores associados a propriedades como *dc:relation* ou *dc:identifier* (para os quais as boas práticas recomendam a utilização de identificadores). Uma solução simples para este problema é usar, de forma complementar ou não, links para termos de vocabulários controlados codificados em linguagens apropriadas (SKOS, RDFS, OWL).

Outro problema comum à maior parte destes repositórios é o fato de todos os recursos serem descritos utilizando o mesmo conjunto de propriedades (as do OAI-DC). Na verdade, há propriedades que são específicas para determinados tipos de recursos. Mais do que isso, recursos do mesmo tipo podem necessitar de propriedades específicas de acordo com o seu domínio de conhecimento (ANDRADE; BAPTISTA, 2014).

Estes são alguns dos problemas facilmente identificáveis em breves análises aos registros de metadados dos repositórios digitais atuais. Os registros OAI-DC são facilmente convertíveis para registros RDF, mas só valeria a pena fazê-lo se esses registros RDF acrescentassem algo ao que atualmente dispomos nos registros OAI-DC. Para isso, esses registros RDF devem estar ligados entre si e ligados a outros registros RDF de natureza diferente (por exemplo, vocabulários controlados, ontologias, conjuntos de dados governamentais, conjuntos de dados estatísticos, dentre outros), ou seja, devem veicular dados cinco estrelas.

A Figura 3 mostra a *LOD cloud*, uma parte da web de dados que cumpre os princípios *Linked Data* (ABELE et al., 2017; BERNERS-LEE, 2009a), isto é, que são dados cinco estrelas (WORLD WIDE WEB CONSORTIUM, 2013). Na *LOD cloud* os dados estão classificados em diversas categorias, e a cada qual é atribuída uma cor. O gráfico é interativo e é possível obter mais informação sobre cada conjunto de dados, bem como verificar os tipos de ligações entre eles. Estes dados, mesmo tendo diferentes naturezas, estão no nível de interoperabilidade mais alta, o que implica que sejam passíveis de relacionamento se não automático, pelo menos com pouca intervenção humana. Quanto mais informação de contexto houver e quanto mais normalizada ela seja, menos intervenção humana se espera.

Um olhar mais aproximado sobre a *LOD cloud* mostra que as redes de repositórios digitais baseadas apenas no protocolo OAI-PMH não estão lá presentes. O OAI-PMH garante níveis de interoperabilidade sintática e uma interoperabilidade semântica básica mas, por si só, não é adequado quando se pretende atingir níveis de interoperabilidade semântica superiores. Passada já quase uma década e meia da criação dos primeiros repositórios digitais que implementam o protocolo OAI-PMH, é surpreendente, e até mesmo desanima-



dor, constatar que estão praticamente no mesmo ponto em termos de interoperabilidade semântica: permanecem em ilhas de interoperabilidade (BAPTISTA, 2010), isto é, são algo interoperáveis entre si, mas estão isolados do resto dos dados e aplicações na web.

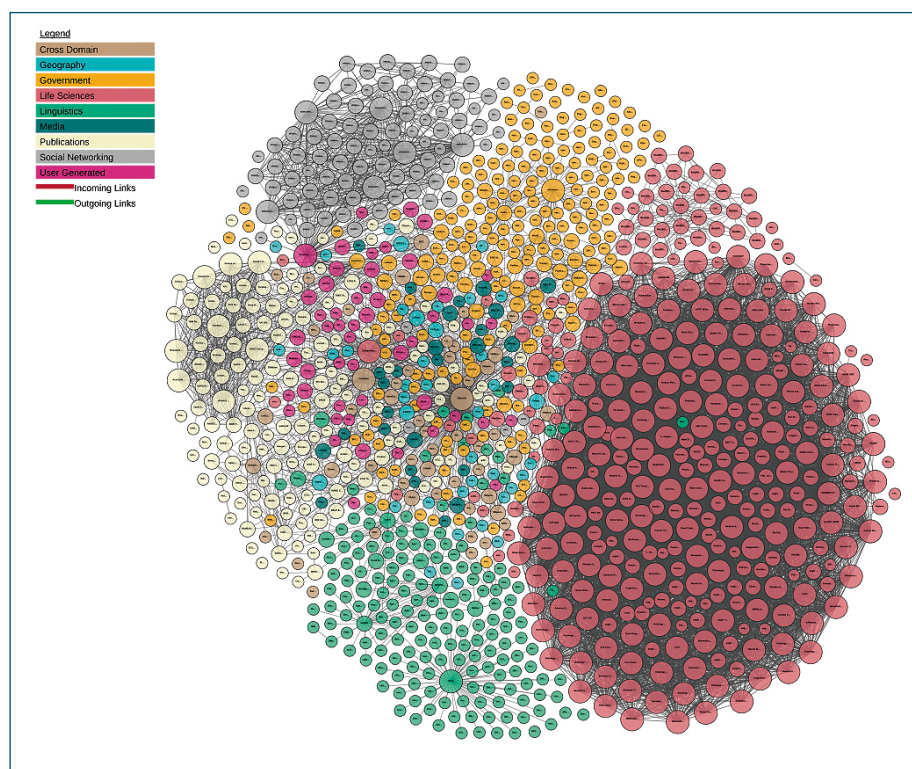


Figura 3 – *Linking open data cloud diagram 2017*

Fonte: Abele et al. (2017).

## PROPOSTA

A posição defendida neste capítulo é, por isso, a seguinte: os repositórios digitais devem estar no nível de interoperabilidade quatro da DCMI e, para isso, devem ter perfis de aplicação rigorosamente definidos, os seus registros de metadados devem estar codificados em RDF e devem estar relacionados com outros dados na web (vocabulários, ontologias, outros dados) que lhes possam fornecer contexto (semântica).

Assim, propõe-se à comunidade Ibero-Americana de metadados em repositórios digitais que desenvolvam e implementem perfis de aplicação observando as seguintes sugestões:



- a) determinar quais propriedades utilizar de acordo com o tipo e a natureza dos recursos. É de esperar que haja várias propriedades em comum que devem ser usadas em recursos de diferentes tipos e naturezas (como, por exemplo, as propriedades do DC Terms), mas também há propriedades que são específicas a determinados tipos de recursos. Por exemplo, supervisor, para teses e dissertações. Os implementadores poderão, ainda, querer incluir propriedades específicas relacionadas com a natureza do recurso. Por exemplo, determinados artigos no âmbito da Medicina poderiam ter associadas propriedades que permitissem informar sobre o diagnóstico ou sobre a terapêutica;
- b) garantir que a semântica das propriedades é rigorosamente definida e aplicada. Se tiverem necessidade de utilizar propriedades com semântica diferente das que têm disponíveis, devem, em primeiro lugar, verificar se as propriedades já estão definidas em algum vocabulário/esquema e se são usadas pela comunidade (utilizando serviços específicos como, por exemplo, o LOV). Se as propriedades não estiverem definidas/declaradas em nenhum vocabulário/esquema, então elas podem ser criadas e disponibilizadas num vocabulário/esquema para que outros as possam também usar. Neste caso, devem estar relacionadas com outras propriedades que já existam, e que lhes fornecem, assim, contexto. Devem posteriormente ser incluídas em serviços como o LOV para que possam ser facilmente encontradas por outros que as queiram utilizar;
- c) garantir que usam refinamentos de propriedades sempre que tal for imprescindível. Esta recomendação está relacionada com a anterior, mas refere-se especificamente às propriedades cuja semântica refina a de outras já existentes. Um caso frequente é o das datas. Existem vários tipos de datas como, por exemplo, *dct:modified*, *dct:valid*, ou *dct:created*. Todas têm uma semântica muito específica e, por vezes, oposta, apesar de todas se referirem a datas. Faz, por isso, sentido que os refinamentos sejam usados e não a propriedade mais geral, neste caso o *dct:date*. Antes de se criar qualquer propriedade nova, deve-se pensar muito bem, já que se trata de um equilíbrio difícil entre rigor semântico e interoperabilidade. A utilização de qualquer nova propriedade, ainda que semanticamente mais rigorosa, pode implicar perda de interoperabilidade;
- d) utilizar esquemas de sintaxe para especificar como se devem interpretar os valores associados às propriedades. Um exemplo frequentemente utilizado é o da indicação da norma ISO 8601 para a representação das datas;
- e) utilizar esquemas de vocabulário para relacionar com os valores de determinadas propriedades como, por exemplo, *dct:subject*, *dct:type* ou *dct:rights*. Se não existirem esquemas de vocabulário que possam usar,

os implementadores podem considerar criá-los e disponibilizá-los livremente à comunidade para uso e reuso. Os registros de metadados podem, assim, ter informação legível por humanos, com as etiquetas dos termos, e informação legível por máquinas com os links diretos para os termos.

- f) utilizar, sempre que possível, informação legível por máquinas (*links*) complementada com informação legível por humanos (texto) para os valores associados às propriedades. Por exemplo, para os valores relacionados com a propriedade `dct:publisher` em vez de se utilizar apenas o nome de determinada editora, pode-se complementar essa informação com o link relativo a essa editora;
- g) utilizar *links* persistentes, ou seja, que permaneçam no tempo.

Estas são as medidas principais a tomar para que se atinjam os níveis de interoperabilidade semântica necessários à incorporação na *LOD cloud* dos dados constantes nos registos OAI-PMH dos repositórios digitais. Para que estes dados passem realmente a fazer parte da *LOD cloud*, um outro passo será ainda necessário: a transformação dos registos OAI-PMH em RDF e a sua disponibilização aberta na rede como *Linked Open Data*.

## CONSIDERAÇÕES FINAIS

Os repositórios digitais organizados em rede são sintaticamente interoperáveis entre si, mas atualmente ainda subsistem muitos problemas de interoperabilidade semântica. Estes devem-se à tecnologia utilizada mas também à qualidade dos metadados: nem dentro das suas próprias ilhas de interoperabilidade os repositórios digitais são semanticamente interoperáveis.

As soluções para estes problemas passam por começar a utilizar tecnologias da Web Semântica, como as baseadas em RDF e repensar a estrutura de metadados dos diversos repositórios digitais, utilizando e adaptando propriedades de diferentes esquemas adequadas às necessidades de cada repositório. Em poucas palavras, criar perfis de aplicação que estejam de acordo com as necessidades específicas dos repositórios.

A comunidade Ibero-Americana de metadados em repositórios digitais é uma comunidade muito dinâmica e empenhada, com vários casos de sucesso a nível mundial. Por isso, é uma comunidade que tem condições para liderar uma mudança no panorama das políticas de interoperabilidade dos repositórios digitais, para dar o passo em frente, evoluindo para o paradigma *Linked Open Data* de maneira calculada e sustentada. Este é o desafio que proponho e para o qual estou disposta a contribuir.

## REFERÊNCIAS

- ABELE, A. et al. **The linking open data cloud diagram**. 2017. Disponível em: <<http://lod-cloud.net/>>. Acesso em: 27 maio 2017.
- BAPTISTA, A. A. A falar nos entendemos: a interoperabilidade entre repositórios digitais. In: GOMES, M. J.; ROSA, F. (Ed.). **Repositórios institucionais: democratizando o acesso ao conhecimento**. Salvador: EDUFBA, 2010. p. 71-90. Disponível em: <<https://repositorio.ufba.br/ri/bitstream/ri/616/3/Repositorios%20institucionais.pdf>>. Acesso em: 27 maio 2017.
- ANDRADE, M. C.; BAPTISTA, A. A. Information needs of researchers in a bibliographic databases environment: a literature review. In: POLYDORATOU, P.; DOBREVA, M. (Ed.). **Let's put data to use: digital scholarship for the next generation: proceedings of the 18th International Conference on Electronic Publishing**. Tassaloniki: IOS Press, 2014. p. 30-38. Disponível em: <<https://doi.org/10.3233/978-1-61499-409-1-30>>. Acesso em: 31 maio 2017.
- ANDRADE, M.; BAPTISTA, A. A. The use of application profiles and metadata schemas by digital repositories: results from a Survey. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2015, São Paulo. **Proceedings...** São Paulo: Dublin Core Metadata Initiative, 2015. p. 146-157. Disponível em: <<http://dcevents.dublincore.org/IntConf/dc-2015/paper/view/362/373>>. Acesso em: 27 maio 2017.
- BERNERS-LEE, T. **Linked data: design issues**. 2009a. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 27 maio 2017.
- BERNERS-LEE, T. **The next web**. 2009b. Disponível em: <[http://www.ted.com/index.php/talks/tim\\_berners\\_lee\\_on\\_the\\_next\\_web.html](http://www.ted.com/index.php/talks/tim_berners_lee_on_the_next_web.html)>. Acesso em: 31 maio 2017. 1 vídeo (17min).
- COYLE, K.; BAKER, T. **Guidelines for Dublin Core application profiles**. 2009. Disponível em: <<http://dublincore.org/documents/2009/05/18/profile-guidelines/>>. Acesso em: 31 maio 2017.
- DUBLIN CORE METADATA INITIATIVE. **DCMI Metadata Terms**. 2012. Disponível em: <<http://dublincore.org/documents/dcmi-terms/>>. Acesso em: 31 maio 2017.
- DUBLIN CORE METADATA INITIATIVE. **Dublin Core metadata element set, version 1.1**. 2008. Disponível em: <<http://dublincore.org/documents/dces/>>. Acesso em: 31 maio 2017.
- HARNAD, S. **Green road to open access: a leveraged transition**. 2007. Disponível em: <<http://users.ecs.soton.ac.uk/harnad/Temp/greenroad.html>>. Acesso em: 31 maio 2017.
- HAUSENBLAS, M. **5\* open data**. 2012. Disponível em: <<http://5stardata.info/en/>>. Acesso em: 30 maio 2017.
- HEERY, R.; PATEL, M. Application profiles: mixing and matching metadata schemas. **Ariadne**, v. 25, set. 2000. Disponível em: <<http://www.ariadne.ac.uk/issue25/app-profiles/>>. Acesso em: 31 maio 2017.
- LINKED Open Vocabularies (LOV). Disponível em: <<http://lov.okfn.org/dataset/lov/>>. Acesso em: 30 maio 2017.
- MILLER, P. **Transcript: sir Tim Berners-Lee talks with Talis about the semantic web**. 2008. Disponível em: <[http://talis-podcasts.s3.amazonaws.com/twt20080207\\_TimBL.html](http://talis-podcasts.s3.amazonaws.com/twt20080207_TimBL.html)>. Acesso em: 31 maio 2017.

NILSSON, M. **Description set profiles**: a constraint language for Dublin Core application profiles. 2008. Disponível em: <<http://dublincore.org/documents/2008/03/31/dc-dsp/>>. Acesso em: 31 maio 2017.

NILSSON, M.; BAKER, T.; JOHNSTON, P. **The Singapore Framework for Dublin Core application profiles**. 2008. Disponível em: <<http://dublincore.org/documents/singapore-framework/>>. Acesso em: 30 maio 2017.

NILSSON, M.; BAKER, T.; JOHNSTON, P. **Interoperability levels for Dublin Core metadata**. 2009. Disponível em: <<http://dublincore.org/documents/interoperability-levels/>>. Acesso em: 31 maio 2017.

WORLD WIDE WEB CONSORTIUM. **W3C semantic web activity**. 2013a. Disponível: <<http://www.w3.org/2001/sw/>>. Acesso em: 31 maio 2017.

WORLD WIDE WEB CONSORTIUM. **5 star linked data**. 2013b. Disponível em: <[https://www.w3.org/2011/gld/wiki/5\\_Star\\_Linked\\_Data](https://www.w3.org/2011/gld/wiki/5_Star_Linked_Data)>. Acesso em: 30 maio 2017.

# **METADADOS E CURADORIA DIGITAL: UM PERFIL DE APLICAÇÃO PARA DESCRIÇÃO DE DADOS DE INVESTIGAÇÃO NA ÁREA DA OCEANOGRAFIA BIOLÓGICA**

Rúbia Tatiana Gattelli





## INTRODUÇÃO

Dados de investigação, dados de pesquisa e dados científicos são os termos em língua portuguesa derivados do inglês *research data* para designar os dados produzidos nas atividades de investigação. Ainda que não haja uniformidade no tocante à nomenclatura adotada para os dados nos países falantes de língua portuguesa, o mesmo não se pode dizer em relação à sua importância para as comunidades acadêmicas e científicas. Se até há alguns anos os dados produzidos nas atividades científicas possuíam apenas valor local, para validar os resultados das pesquisas que os originaram, hoje a consciência de sua importância se expande a cada dia.

O paradigma científico conhecido como eScience trouxe embutido em suas características um novo olhar para os dados de investigação. Impulsionada pelo avanço das tecnologias de informação e comunicação, a eScience pode ser definida como uma metodologia de investigação (GORE, 2011; PENNOCK, 2006; WHITMIRE, 2013) que utiliza largamente os recursos tecnológicos disponíveis para empreender as atividades científicas. Tal tecnologia alterou as características do fazer científico, melhorando e aumentando a capacidade de produção de dados, assim como alterou a relação entre os investigadores, modificando e estimulando a comunicação científica. De fato, para Gore (2011) duas das principais características que definem a e-Science são os dados produzidos nas atividades de pesquisa e o trabalho colaborativo.

Adicionalmente, Osswald (2008) declara que a *eScience* é baseada em redes distribuídas de computadores e softwares que fornecem a capacidade necessária para processar grandes conjuntos de dados. Para o autor, esse potencial tecnológico também traz consigo um componente comportamental, pois estimula um aumento de troca de informações e intensifica a comunicação entre os investigadores, melhorando assim sua competitividade e sua habilidade de cooperação, seja em nível local ou internacional.

Se, por um lado, o aporte tecnológico trouxe benefícios e facilidades para a produção e a comunicação de dados, em contrapartida trouxe novos desafios para as comunidades científicas. Um deles é o de lidar com volumes cada vez maiores de dados, que podem ser produzidos em larga escala. E outro é a necessidade de assegurar que estes dados estejam armazenados de forma segura e com sua qualidade preservada ao longo do tempo.



Além de se ocupar com a produção dos dados durante a prática científica, os cientistas precisam ao mesmo tempo se preocupar com a sua salvaguarda. A gestão de dados de investigação passou, portanto, a ser considerada tópico de absoluta relevância para vários grupos ligados à ciência. Uma gestão adequada permite que dados possam ser solicitados a qualquer momento para comprovar a veracidade de resultados de pesquisas. Os dados de investigação produzidos no âmbito das pesquisas fomentadas por verbas públicas ganham especial visibilidade e devem, portanto, estar disponíveis para o escrutínio de toda a sociedade.

Além de conferir transparência às atividades de investigação, as boas práticas de gestão asseguram que os dados estejam acessíveis e inteligíveis para sua reutilização em novos projetos de pesquisa, proporcionando maior colaboração entre cientistas e economia na coleta de novos dados. Essas práticas podem incluir planos de gestão de dados, políticas institucionais de uso e partilha de dados, treinamentos necessários (literacia para o uso e reuso de dados), procedimentos para o armazenamento seguro dos dados, barreiras de acesso e proteção aos conjuntos de dados, documentação dos dados, entre muitas outras práticas.

Para Corti et al. (2014), a responsabilidade dos pesquisadores em relação aos seus dados está mudando. Estes precisam melhorar, aumentar e profissionalizar suas habilidades para a gestão de dados, de modo que possam enfrentar, de forma responsável e eficiente, o desafio da produção de resultados de pesquisa de alta qualidade e de dados sustentáveis, habilidades cujo desenvolvimento lhes permitirá a aptidão para o compartilhamento e reutilização dos seus dados e, por consequência, dos seus resultados de pesquisa.

## **DADOS DE INVESTIGAÇÃO: CONCEITO E CURADORIA DIGITAL**

Os dados de investigação são estruturas de definição complexa, pois podem ter significados diferentes para diferentes pesquisadores ou em diferentes contextos. Há que se levar em conta ainda a linguagem técnica usada nos variados ambientes de investigação, o que pode conferir diversidade semântica relativamente ao conceito de dados.

O MANTRA<sup>1</sup>, curso online sobre gestão de dados de investigação mantido por funcionários dos Serviços de Informação da Universidade de Edimburgo, na Escócia, define dados como elementos de baixo nível de abstração, a partir dos quais a informação e o conhecimento são derivados; são recolhidos, observados ou criados para fins de análise com o objetivo de produzir e validar resultados de investigações. Em outras palavras, dados são produzidos no nível mais primá-

<sup>1</sup> Disponível em: <<http://mantra.edina.ac.uk/researchdataexplained/>>. Acesso em: 16 out. 2014.



rio da prática investigativa, nível no qual medidas são originalmente coletadas, como, por exemplo, medidas de temperatura da água, registro da velocidade e direção do vento, imagens capturadas de satélite, respostas individuais a um inquérito ou censo, dentre outros. Esses tipos de dados podem ser considerados em alguns grupos como dados brutos e podem ser manipulados de diversas formas gerando, assim, outros tipos de dados mais refinados.

Surgem na literatura diversas classificações para os dados quanto à sua origem e tipologia. Segundo Sayão e Sales (2015), dados podem ser originados pela observação direta de um evento, o que os associa a tempos e locais específicos (dados observacionais); também podem ser resultado de execução e simulação de modelos computacionais (dados computacionais), ou provenientes de experimentos de laboratório (dados experimentais). O MANTRA acrescenta ainda duas classificações de dados quanto à sua origem: os dados derivados ou compilados, resultantes do processamento ou combinação de dados brutos, e os dados de referência ou canônicos, que são a conglomeração ou coleção de conjuntos de dados menores, muito provavelmente publicados e curados.

Quanto à tipologia, o MANTRA afirma que os dados podem variar quanto ao formato físico e tipo de conteúdo, podendo ser classificados em dados textuais, numéricos, multimídia, modelos, softwares, dados específicos de disciplinas e dados específicos de instrumentos. Podem ser considerados objetos de dados documentos de texto (Word), planilhas (Excel), notas de campo e laboratório, material de áudio e vídeo, fotografias, filmes, artefatos, espécimes, amostras, algoritmos, scripts, workflows, procedimentos padrão e protocolos, entre muitos outros.

Tais definições são amplas o suficiente para considerar a importância dos dados produzidos nas mais variadas disciplinas, não somente como resultados de investigação, mas como capital para novos empreendimentos científicos. Compreender a natureza e tipologia dos dados é importante para que sejam desenvolvidos planos de gestão de acordo com suas características e necessidades.

Dados são considerados, portanto, o insumo básico de qualquer projeto de investigação, e o seu ciclo de vida está intimamente ligado ao ciclo de vida do projeto. No entanto, há que se observar a validade dos dados de investigação para além da validade do seu projeto, uma vez que este pode ser encerrado, mas os dados podem manter um valor ativo para futuras práticas investigativas, iniciando um novo ciclo de vida. Contudo, para que isso seja possível, é necessário que os dados tenham sido bem geridos durante o projeto de investigação, é necessário que estejam curados para manter sua qualidade preservada e seu acesso garantido para reutilização.

O sentido de curadoria para dados de investigação é o mesmo da curadoria de coleções já tradicionalmente conhecidas, e está relacionado com cuidado

para preservação ao longo do tempo, prevendo aspectos de descrição e contextualização (a exemplo do que vemos nas coleções de museus). O que difere é que, assumindo-se que hoje a maioria dos conjuntos de dados recolhidos ou gerados nas atividades de investigação já estão no formato digital, produzidos neste formato ou digitalizados, a curadoria é essencialmente digital.

A curadoria digital, conforme o *Digital Curation Centre* (DCC)<sup>2</sup>, envolve a manutenção, preservação e adição de valor aos dados de investigação digitais ao longo do seu ciclo de vida. Estas ações evocam a ideia de sustentabilidade dos dados desde a sua produção, recolha, ou mesmo recebimento de conjuntos de dados provenientes de outro projeto, durante toda sua utilização em uma ação investigativa, até seu armazenamento adequado e seguro para uma futura reutilização.

A adição de valor aos dados de investigação é umas das ações essenciais de curadoria digital, que envolve práticas de descrição e contextualização aos dados. Significa fornecer informação extra sobre as características dos dados (físicas, estruturais, conteúdo) e sobre o contexto onde foram criados (motivação, localização temporal e espacial, metodologia utilizada). Esses aspectos são primordiais para que um conjunto de dados seja entendido por terceiros ou mesmo por quem os produziu depois de passado algum tempo.

A boa documentação ou anotação dos dados de investigação, termos também usados na literatura para designar descrição, é fator crítico para sua reutilização e preservação. Para Abbott (2008) os metadados são o tipo de anotação descritiva crucial para recuperação da informação, pois, além de fornecerem a informação necessária para o entendimento dos dados, também fornecem os pontos de acesso para que eles sejam encontrados em uma determinada base de dados. Corti et al. (2014) complementam, afirmando que os metadados fornecem informação estruturada e padronizada que explica a proposta e origem dos dados, cria autoria, condições de acesso e termos de uso da coleção de dados. Qin e D'Ignazio (2010) também concordam que os metadados desempenham um papel central nas atividades relacionadas aos dados em múltiplos aspectos durante o processo de investigação.

De fato, podemos considerar os metadados como elementos que agregam valor aos dados de investigação, os organizam, facilitam sua descoberta e promovem interoperabilidade entre recursos. De igual modo, a adequada gestão de dados depende fortemente de registros precisos de metadados que promovam sua interpretação, reuso e previnam sua perda (CASTRO et al., 2016). Entretanto, metadados provenientes de padrões e esquemas disponíveis podem não ser suficientes para contemplar descritivamente a diversidade de conteúdo e forma que abrangem os dados de investigação.

2 Disponível em: <<http://www.dcc.ac.uk/digital-curation/what-digital-curation>>. Acesso em: 16 abr. 2017.

A apropriada documentação de conjuntos de dados de investigação demanda a adoção de metadados que possam cobrir a extensão de diferentes domínios. No entanto, para grupos de investigação oriundos das universidades, que normalmente não possuem os recursos necessários para investir em curadoria, uma das únicas opções seria o depósito dos seus dados no repositório institucional. O problema é que os repositórios institucionais utilizam um padrão de metadados bastante genérico, abrangente o suficiente para descrever toda a sorte de objetos intelectuais produzidos no âmbito da instituição, o que certamente negligenciaria aspectos importantes na descrição de dados de investigação.

Mesmo os repositórios de dados podem não possuir a capacidade semântica adequada para representar as especificidades dos diferentes conjuntos de dados. Segundo White (2014, p. 24):

Os repositórios de dados científicos apresentam um ambiente único onde a organização do conhecimento e o gerenciamento de informações pessoais convergem. Os profissionais de informação criam repositórios baseados em teorias e esquemas de organização de conhecimentos tradicionais que têm sido utilizados pelas bibliotecas por anos. No entanto, os conjuntos de dados criam um desafio único. Um conjunto de dados de investigação é criado por um investigador ou grupo de investigadores e é organizado de uma forma que reflete as suas necessidades.

Logo, se um repositório de dados deve refletir as necessidades de descrição e organização da informação dos cientistas, deve ser desejável que também considere a pluralidade destes grupos de investigação, das atividades que desenvolvem e dos dados que produzem. Diferentes domínios demandam conjuntos específicos de metadados que descrevam seus dados de investigação de forma eficiente, evitando o risco de que percam sentido ao longo do tempo. Dados de um domínio específico dentro das ciências humanas, por exemplo, podem ser absolutamente diferentes dos dados produzidos no âmbito das disciplinas das ciências exatas, e, portanto, devem exigir um rigor descritivo também diferenciado.

A necessidade de descrever conjuntos de dados de diferentes domínios tem levado à criação de perfis de aplicação (CASTRO; RIBEIRO; SILVA, 2013). Para Heery e Patel (2000) os perfis de aplicação (do inglês *application profile*) são um tipo de esquema de metadados que consistem em elementos de dados retirados de um ou mais esquemas existentes, combinados entre si e otimizados para uma aplicação local em particular. Adicionalmente, temos a definição de Duval et. al. (2002):

Um perfil de aplicação é um conjunto de elementos de metadados selecionados de um ou mais esquemas de metadados e combinados em um esquema composto. Perfis de aplicação fornecem os meios para expressar princípios

de modularidade e extensibilidade. O propósito de um perfil de aplicação é adaptar ou combinar esquemas existentes em um pacote adaptado aos requisitos funcionais de um aplicativo específico, mantendo a interoperabilidade com os esquemas de base originais. Parte dessa adaptação pode incluir a elaboração de elementos de metadados locais que tenham importância em determinada comunidade ou organização, mas que não se espera que sejam importantes em um contexto mais amplo.

Conforme as definições acima, podemos reconhecer os perfis de aplicação como ferramentas importantes na curadoria de dados de investigação, uma vez que estes podem ser desenhados sob medida para a descrição de domínios específicos. Além de contemplar metadados descritivos, os perfis de aplicação implicam também no armazenamento seguro dos dados, pois são pensados para aplicações específicas.

A seguir, este texto abordará a metodologia utilizada para desenhar um perfil de aplicação para investigadores do domínio da Oceanografia Biológica provenientes da Universidade Federal do Rio Grande (FURG). O perfil foi desenhado como parte da dissertação de mestrado da autora deste capítulo e faz parte de um trabalho conjunto realizado no *Information Systems Laboratory* (InfoLab), da Faculdade de Engenharia da Universidade do Porto (FEUP), em Portugal. A equipe do InfoLab trabalha em conjunto com um painel de investigadores de diversos domínios na elaboração de modelos específicos de metadados. Os descritores destes modelos específicos de cada domínio são formalizados em ontologias constituindo vocabulários que são incorporados em uma plataforma colaborativa e multidisciplinar de gestão de dados de investigação.

A plataforma, que recebeu o nome de Dendro, também foi desenvolvida no âmbito do trabalho que é realizado no InfoLab/FEUP. É uma plataforma colaborativa para organização e descrição de dados de investigação, onde os investigadores são encorajados a depositar e descrever seus conjuntos de dados de forma simples e customizada, utilizando os vocabulários disponíveis. A plataforma Dendro ainda garante interoperabilidade com plataformas externas.

## **PERFIL DE APLICAÇÃO *BIOLOGICAL OCEANOGRAPHY***

O perfil de aplicação, denominado *Biological Oceanography*, foi desenhado com a colaboração de investigadores de três laboratórios do Núcleo da Oceanografia Biológica da FURG, sendo estes: Laboratório de Crustáceos Decápodes, Laboratório de Ictiologia e Laboratório de Ecologia de Invertebrados Bentônicos. O perfil de aplicação busca, portanto, retratar a prática investigativa e a necessi-

dade de descrição dos dados gerados nestas subáreas, sem a intenção de abranger a completude do domínio.

Para o melhor entendimento do perfil de aplicação, segue uma explanação sobre o contexto organizacional da pesquisa científica no domínio da Oceanografia Biológica, bem como uma descrição das atividades prático-científicas realizadas na área.

## A UNIVERSIDADE: CONTEXTO ORGANIZACIONAL

A FURG é uma instituição pública de ensino superior, situada na região sul do Brasil, com sede na cidade de Rio Grande, estado do Rio Grande do Sul. Nasceu na década de 1950, quando foi estrategicamente instituída a Escola de Engenharia. Na época não havia escolas de nível superior na cidade de Rio Grande, e a área de engenharia apresentava uma grande demanda de profissionais. Aos poucos surgiram outras unidades educacionais, que aglutinaram-se e, em 1969, formaram oficialmente a então nomeada Universidade do Rio Grande (URG). Em 1987, passa à condição de Fundação Pública, sendo custeada substancialmente por recursos da União Federal.

O ano de 1987 também marca a definição da Filosofia e Política da Universidade do Rio Grande, onde assume como vocação institucional o Ecossistema Costeiro, que orientará as atividades de ensino, pesquisa e extensão. Hoje a Universidade tem grande importância no cenário educacional nacional, sobretudo no que se refere ao oferecimento de formação em cursos ligados ao meio ambiente marinho e costeiro.

Através do Instituto de Oceanografia, a FURG oferece hoje cursos em nível de graduação e pós-graduação voltados para o estudo e a investigação dos ecossistemas costeiros. A oferta educativa inclui dois cursos de graduação: Oceanologia (criado em 1970, o primeiro do Brasil) e Tecnologia em Gestão Ambiental (oferecido nas cidades de Rio Grande e São Lourenço do Sul). A oferta formativa em nível de pós-graduação inclui um Programa de Pós-Graduação em Oceanografia Biológica (com cursos de mestrado e doutorado), um Programa de Pós-Graduação em Oceanografia Física, Química e Geológica (com cursos de mestrado e doutorado), um Programa de Pós-Graduação em Aquicultura (com cursos de mestrado e doutorado), um Programa de Pós-Graduação em Gerenciamento Costeiro (com curso de mestrado).

A FURG também apresenta a infraestrutura necessária para confirmar sua prevalência na área de investigação costeira. No Câmpus Carreiros está consolidada a Base Oceanográfica, que possui vários laboratórios de ensino e pesquisa, a Biblioteca Setorial de Pós-Graduação em Oceanografia, que possui acervo es-

pecializado na área, além da Estação de Apoio Antártico (ESANTAR), responsável pela manutenção e guarda de todo o material utilizado na Operações Antárticas. No centro da cidade, a Universidade conta com o Complexo de Museus e Centros Associados, formado pelo Museu Oceanográfico Eliézer de Carvalho Rios, Museu Antártico, Eco-Museu da Ilha da Pólvora, Museu Náutico, Centro de Recuperação de Animais Marinhos (CRAM) e Centro de Convívio dos Meninos do Mar (CCMar). No Balneário Cassino está localizada a Estação Marinha de Aquacultura Prof. Marcos A. Marchiori (EMA).

A FURG ainda conta com uma frota de veículos tracionados utilizados nas saídas de campo, e uma frota flutuante, que compreende o navio oceanográfico Atlântico Sul, a lancha oceanográfica Larus e o bote Prof. Moralles, embarcações imprescindíveis para o desenvolvimento das atividades de investigação e a formação dos oceanólogos. Também conta com o apoio de navios da Marinha do Brasil, que podem ser partilhados para estudos costeiros e oceânicos.

Além de tudo, a FURG faz parte de vários programas, projetos e comissões voltados para o ensino, investigação, extensão e políticas públicas que envolvem as ciências do mar, com atuação expressiva no cenário nacional e no exterior. Sendo assim, a FURG historicamente consolida sua importância para o ensino, a pesquisa e a extensão na área de ciências do mar. A produção científica gerada especialmente em nível de pós-graduação projeta a universidade no panorama científico nacional e internacional e promove o desenvolvimento regional. Os dados resultantes desta produção científica possuem valor de curto e longo prazo e, portanto, devem ser objetos de curadoria.

### A Área de Investigação da Oceanografia Biológica na FURG

O Programa de Pós-Graduação em Oceanografia Biológica oferece cursos de Mestrado e de Doutorado nas seguintes linhas de pesquisa: Oceanografia Pesqueira; Biotecnologia Marinha e Oceanografia Aplicada; Poluição, Modelagem e Gerenciamento Ambiental; e Sistemática, Biologia e Ecologia de Organismos Marinhos e Estuarinos. Para o desenvolvimento das linhas de pesquisa, conta com o suporte de nove laboratórios de subáreas específicas da Oceanografia Biológica.

As pesquisas estão voltadas ao estudo das relações ecológicas entre os componentes bióticos e abióticos dos ambientes costeiro e marinho, à oceanografia pesqueira, estudos de poluição marinha, modelagem ecológica e gerenciamento ambiental, bem como oceanografia aplicada e biotecnologia marinha. Os estudos desenvolvem-se numa ampla variedade de ambientes como, por exemplo, estuários, águas costeiras e mar profundo do sudoeste do Oceano Atlântico, abrangendo toda Zona Econômica Exclusiva brasileira, ilhas oceânicas tropicais e região Antártica.

Devido à determinante geográfica da cidade de Rio Grande, grande parte dos estudos do programa de pós-graduação em oceanografia biológica se desenvolve no estuário da Lagoa dos Patos, que margeia a cidade. O estuário possui importância ecológica, econômica e social para as comunidades que vivem em seus arredores, onde são encontrados muitos pescadores artesanais, algumas indústrias de pescados e um potente polo industrial.

## CARACTERIZAÇÃO DA PRÁTICA CIENTÍFICA E PRODUÇÃO DE DADOS NO DOMÍNIO DA OCEANOGRAFIA BIOLÓGICA

Foram abordadas neste estudo a prática científica e a produção de dados de investigação em três subáreas do domínio da Oceanografia Biológica, nomeadamente nas áreas de ecologia de peixes, ecologia de crustáceos decápodes e ecologia de organismos invertebrados bentônicos. O foco de estudo das áreas é investigar as interações ecológicas entre organismos marinhos e estuarinos e os ambientes onde vivem.

Os dados de investigação são gerados basicamente a partir de duas práticas distintas que foram designadas como atividades de campo e atividades de laboratório. As atividades de campo dizem respeito aos eventos de coleta de material biológico e ao registro dos parâmetros ambientais e são realizadas normalmente no meio ambiente estuarino e no meio ambiente praial. Esses eventos geram uma série de dados observacionais, que são capturados em tempo real. As atividades realizadas em laboratório dizem respeito aos tipos de interações necessárias realizadas com o material biológico coletado em campo. Essas atividades podem envolver a tomada de diversas medidas ou envolver algum tipo de experimento, gerando dados dos tipos observacionais e experimentais.

Os principais parâmetros ambientais registrados para as atividades dos três grupos de investigadores são salinidade, transparência e temperatura da água, profundidade da coluna d'água, topografia do fundo do estuário; eventualmente podem ser registrados velocidade e direção do vento, altura de onda e direção da corrente (no caso de coleta na praia). Dados de precipitação e vazão do estuário podem ser necessários e são consultados em bancos de dados especializados. Os parâmetros ambientais, chamados pelos investigadores de dados abióticos, podem ser registrados durante os eventos de coleta de material biológico ou, independentemente, de acordo com periodicidade previamente estipulada.

O material biológico coletado, composto por organismos e sedimentos, é levado para os laboratórios, onde inicialmente é realizada a triagem. Na triagem dos sedimentos, estes são separados, e calculados os elementos que os compõem, tais como argila, areia, silte e matéria orgânica. Na triagem dos orga-

nismos, os indivíduos são separados por espécies e, então, são feitas uma série de contagens (quantidades de espécies, de indivíduos por espécie e total de indivíduos capturados) e tomadas uma série de medidas (biometrias dos indivíduos, que são diferentes para cada espécie, mas que consistem basicamente em medidas de largura, comprimento e peso). Ainda podem ser verificadas outras variáveis, como o sexo, estágio de vida e o estágio de muda (para crustáceos) dos indivíduos. Os dados derivados das interações com os organismos são designados pelos investigadores como dados bióticos.

Para ambos os dados, bióticos e abióticos, são definidos e registrados alguns dados espaciais e alguns dados temporais, necessários para contextualizar as atividades de coleta territorial (como: nome do local onde será realizado o evento, nomes dos pontos específicos de coleta daquele local, suas coordenadas geográficas, a quantidade de pontos de coleta e a extensão do perfil) e temporal (a exemplo de: data do evento, período de abrangência dos eventos, a periodicidade em que se realizam os eventos e a estação do ano em que se realizam.). Cada uma das atividades (de campo e de laboratório) também lida com metodologia, equipamentos e instrumental específicos.

O registro dos dados coletados é sempre feito em duas etapas. Primeiramente, são registrados à mão em planilhas de papel no momento de sua coleta; depois, são transcritos em planilhas eletrônicas, do tipo Excel. Sua transposição das planilhas de papel para as planilhas eletrônicas é uma operação delicada e pode resultar em erros de digitação caso não seja feita com atenção. As planilhas eletrônicas são armazenadas em pastas nos computadores dos laboratórios e organizadas por nomes e datas. Nem sempre são feitas cópias de segurança ou armazenamento nas nuvens, ficando a critério de cada investigador.

Os dados bióticos e abióticos, considerados pelos investigadores dados brutos, depois de passarem por análises estatísticas geram outros tipos de dados. Estes levam os investigadores ao objetivo final de suas pesquisas, resultando em teses, relatórios e artigos científicos, os quais são objetos de comunicação científica. Existe um sentido de cuidado para a preservação dos dados brutos, visto que os investigadores muitas vezes reutilizam os conjuntos de dados, normalmente com investigadores e estudantes de pós-graduação do mesmo laboratório, ou dos laboratórios vizinhos; no entanto, ainda não está estabelecida a prática de comunicar os dados brutos e de disponibilizá-los para reuso de outros. Portanto, questões como gestão e partilha de dados são realizadas informalmente, dependendo de iniciativas particulares.



## Abordagem Metodológica para o Desenvolvimento de Modelos de Metadados de Domínios Específicos: o Caso da Oceanografia Biológica

Modelos de metadados de diversos domínios estão sendo elaborados pelos investigadores do InfoLab/FEUP em conjunto com um painel de investigadores oriundos destes domínios. Tais modelos são incorporados na Plataforma Dendro para gestão de dados de investigação, permitindo aos investigadores depositar e descrever seus dados de forma acurada. Descritores específicos de cada domínio devem ser simples e estar de acordo com a terminologia utilizada pelos investigadores, de modo que estes possam facilmente adoptá-los na documentação de seus conjuntos de dados (CASTRO et al., 2016).

O modelo de metadados elaborado para as experiências de investigação no domínio da Oceanografia Biológica seguiu a metodologia desenvolvida no InfoLab, cujo passo-a-passo pode ser verificado abaixo.

### Passo um: levantamento de requisitos do domínio

Antes mesmo de levantar os requisitos necessários para a modelagem dos descritores específicos do domínio, é imprescindível que os investigadores estejam envolvidos em todo o processo. Sendo os reais conhecedores do domínio, o trabalho conjunto é fundamental. Sugere-se que seja marcado um primeiro encontro para que o curador explique ao investigador o processo como um todo e o papel que ambos devem desempenhar. No caso da experiência com os investigadores da Oceanografia Biológica, o encontro foi coletivo, e foi realizada apresentação sobre boas práticas de gestão de dados de investigação, a fim de sensibilizá-los sobre sua importância.

Em seguida, marcou-se com cada investigador uma entrevista, que consistia em questões mais específicas, sobre o fluxo dos dados produzidos durante o ciclo de investigação, e de questões mais genéricas, sobre atividades de gestão, ao armazenamento e compartilhamento de dados. O roteiro da entrevista foi enviado previamente por e-mail, para que os investigadores pudessem se preparar com antecedência. As perguntas feitas faziam parte do **Guião de Entrevista Curadoria de Dados**, traduzido e adaptado de um instrumento disponibilizado pelo Data Curation Profiles<sup>3</sup>.

A entrevista com cada um dos investigadores foi gravada para ser analisada posteriormente. No momento da entrevista, também solicitou-se a cada investigador uma amostra de um conjunto de dados e dois artigos científicos de sua autoria publicados, dos quais se procedeu a uma análise de conteúdo.

<sup>3</sup> Disponível em: <<http://datacurationprofiles.org/>>. Acesso em: 20 jul. 2017.

A leitura e análise de conteúdo dos artigos, juntamente com a visualização dos dados e a entrevista, possibilitaram maior entendimento de como é feita a prática investigativa na área e de como são gerados os dados. Além disso, ajudou na compreensão da linguagem técnica usada pelos cientistas. Ainda assim, muitas dúvidas surgiram, e precisou-se recorrer a pesquisas na Internet, consulta a dicionários e vocabulários da área, e aos próprios investigadores, estes, sem dúvida, a melhor fonte de informação.

### Passo dois: a modelagem do domínio em mapas conceituais

Depois de levantados os principais requisitos do domínio, foi possível modelar o conhecimento da área estabelecendo relações entre os conceitos, fundamentalmente extraídos da análise de conteúdo das publicações, através do desenho de um mapa conceitual. Com o mapa, pôde-se retratar o fluxo das atividades práticas de investigação das áreas e as variáveis intrínsecas para a produção de dados. Para o desenho do mapa de conceitos, foi utilizado o software *CMapTools*<sup>4</sup>.

Inicialmente, foram desenhados três mapas, um para cada subárea estudada. Depois de validados junto aos investigadores, foram reunidas todas as características comuns que eles apresentavam, e os mapas foram unidos em apenas um, representativo das três áreas, como pode ser visualizado na Figura 1.

A tarefa posterior é identificar quais conceitos do mapa poderão se tornar descritores do modelo de metadados. Essa tarefa exige do curador um bom conhecimento em relação à organização e representação da informação, isso assegura que ele possua a capacidade de depreender do mapa conceitual os descritores capazes de representar todas as facetas desejadas do domínio. Além disso, deve dotar os investigadores da noção de descritor (metadados) e qual sua função na documentação dos seus dados de investigação, para que estes possam atuar efetivamente na escolha dos descritores.

Durante a transposição de conceitos para descritores, observou-se a necessidade de estabelecer alguns objetivos para o modelo de metadados de modo a delimitar e nortear o processo. Os objetivos foram determinados conforme dúvidas e necessidades apontadas pelos investigadores. Eles sentiram falta, por exemplo, de alguns descritores ausentes no mapa de conceitos, descritores que contemplassem aspectos dos projetos de pesquisa. A seguir encontram-se alguns exemplos dos objetivos estipulados:

- a) escolher descritores que cubram o domínio de forma abrangente, porém o mais completa possível;

<sup>4</sup> Disponível em: <<http://cmap.ihmc.us/>>. Acesso em: 20 jul. 2017.

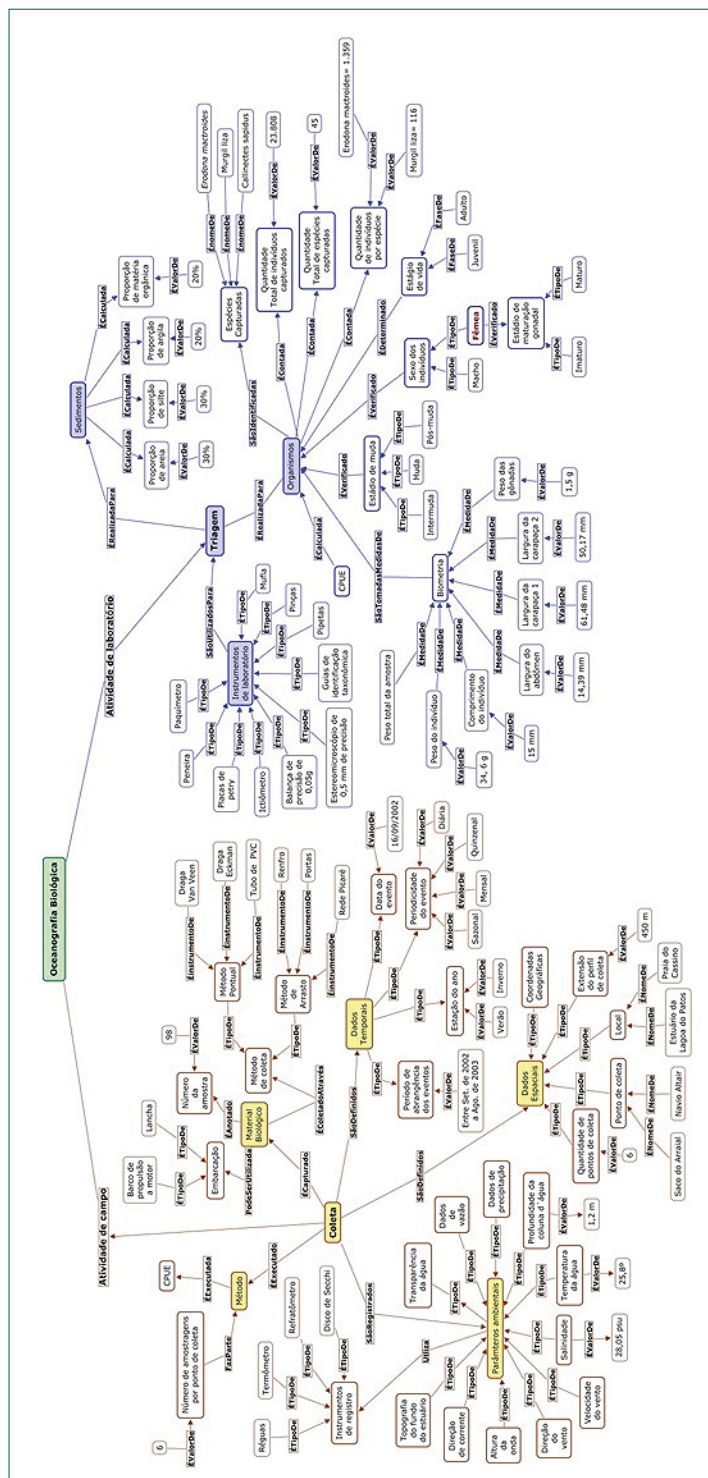


Figura 1 – Mapa de conceitos  
Fonte: Autoria própria (2015).

- b) definir aspectos de ordem administrativa relativos ao projeto, tais como descrição ou resumo do projeto, responsável/chefe pelo laboratório, órgão de financiamento;
- c) definir elementos que identifiquem o conjunto de dados, tais como título, autoria, assunto, natureza ou tipologia dos dados;
- d) identificar material coletado (nome comum, nome científicos);
- e) quantificar espécies e indivíduos coletados (contagem, peso);
- f) localizar os eventos de coleta de dados no tempo (data, periodicidade de coleta) e no espaço (local, coordenadas geográficas);
- g) especificar métodos e instrumentos para amostragens;
- h) especificar destinação de materiais coletados.

### Passo três: estudo de esquemas de metadados e transposição dos descritores escolhidos para metadados extraídos dos esquemas

Concomitantemente com o processo de escolha dos descritores, iniciou-se a busca por esquemas de metadados e normas de descrição de dados existentes no domínio da oceanografia biológica ou em domínios relacionados. A ideia era encontrar nos vocabulários existentes os descritores necessários para descrever o domínio com todas as especificidades que demanda. No sítio eletrônico do *Digital Curation Centre*, há uma seção, organizada por disciplinas, que lista padrões para descrição de dados e casos de usos em vários domínios. Foram escolhidos três padrões da categoria **Biology** (Darwin Core, *Ecological Metadata Language* (EML) e OBIS *Data Schema*), e um padrão da categoria *Earth Science* (*Content Standard for Digital Geospatial Metadata*, Part 1: *Biological Data Profile* (FGDC/CSDGM)). Estes padrões contemplariam descritores bastante específicos do domínio. Ainda, foi escolhido um padrão da categoria **General Research Data** para contemplar os aspectos mais genéricos do domínio, sendo que o escolhido foi o *Dublin Core*.

Depois de selecionados todos os conceitos a partir do mapa conceitual e dos objetivos estabelecidos pelos investigadores, estes foram confrontados com os descritores da lista de esquemas supracitada. O objetivo desta tarefa é encontrar nos vocabulários existentes descritores que correspondam semanticamente ao que os conceitos pretendem representar. Em alguns casos foi possível encontrar compatibilidade semântica em descritores de mais de um dos vocabulários para representar um conceito. Para resolver esta situação, foi necessário estabelecer uma ordem de prioridade para uso dos esquemas, o primeiro dos quais foi o padrão *Dublin Core*, por ser amplamente conhecido e utilizado, garantindo interoperabilidade com outras aplicações e repositórios. Em seguida, vieram os esquemas EML e *Darwin Core*, por sua linguagem compatível com o domínio.

O Quadro 1 apresenta um exemplo de como foi realizada a transposição dos conceitos desejados para descritores dos esquemas de metadados selecionados. Os conceitos estão separados pelos objetivos previamente estipulados, e os descritores apresentam uma breve explanação conceitual.

CONCEITO	DESCRIPTOR SUGERIDO	ESQUEMA	DESCRIÇÃO
<b>OBJETIVO:</b> Definir aspectos de ordem administrativa relativos ao projeto			
Nome do Laboratório /e/ou/ Nome da Instituição	<i>Organization Name</i>	EML	The responsible party field contains the full name of the organization that is associated with the resource. This field is intended to describe which institution or overall organization is associated with the resource being described. Ex.: National Center for Ecological Analysis and Synthesis.
Órgão de financiamento	<i>Funding</i>	EML	The funding field is used to provide information about funding sources for the project such as: grant and contract numbers; names and addresses of funding sources. Other funding-related information may also be included.
<b>OBJETIVO:</b> Definir elementos que identifiquem o conjunto de dados			
Autor (responsável pelo projeto que origina a coleta dos dados)	<i>Creator</i>	<i>Dublin Core</i>	An entity primarily responsible for making the resource.
Contato do autor (e-mail)	<i>Electronic Mail Address</i>	EML	The electronic mail address is the email address for the party. Example(s): my-email@mydomain.edu
Título	<i>Title</i>	<i>Dublin Core</i>	A name given to the resource.
Assunto	<i>Subject</i>	<i>Dublin Core</i>	The topic of the resource. Typically, the subject will be represented using keywords, key phrases, or classification codes.
<b>OBJETIVO:</b> Especificar destinação de materiais coletados			
Destinação da amostra	<i>Sample Destination</i>	<i>Biological Oceanography</i>	Describes the final destination of a sample after used in the research analysis. For e.g.: discarded, destroyed, donated to another research project, deposited in the lab collection, etc.

Quadro 1 – Transposição de conceitos para descritores oriundos de esquemas de metadados

Fonte: Autoria própria (2015).

Ao total, o modelo de metadados específico para as experiências de investigação da Oceanografia Biológica contou com 31 descritores: 8 *Dublin Core*; 12 EML; 5 *Darwin Core*; 1 OBIS; e outros 5 descritores criados sob o prefixo BIOCIN, pois não encontraram compatibilidade semântica com nenhum descritor existente nos esquemas estudados.

#### Passo quatro: representação dos descritores em ontologia e incorporação do perfil à plataforma Dendro

Depois de selecionados todos os descritores necessários para compor o modelo de metadados para o domínio da oceanografia biológica, a próxima etapa é a sua formalização em ontologia para que possam ser inseridos na plataforma Dendro e funcionem como um perfil de aplicação para descrição de dados de investigação do domínio específico.

A ontologia dentro da plataforma funciona como um vocabulário usado para descrever um domínio, onde as propriedades são a representação dos descritores, ou seja, os descritores são representados pelas propriedades da ontologia. É importante que a linguagem da ontologia seja a mesma do modelo de metadados para não haver barreira no entendimento dos descritores. Neste caso, a ontologia constitui-se apenas de algumas classes e propriedades e, por apresentar este formato simplificado, é chamada de **ontologia ligeira** (*lightweight ontology*) (CASTRO; SILVA; RIBEIRO, 2014).

Ao comparar o modelo completo de metadados desenhado para as necessidades de descrição da oceanografia biológica com outros vocabulários já ingestados na plataforma Dendro em forma de ontologia, observou-se a desnecessidade de formalizar todos os descritores. Muitos destes que compuseram o modelo já estavam representados em outras ontologias na plataforma, como, por exemplo, a ontologia *Dublin Core*, portanto, apenas os que não estavam contemplados foram representados em forma de *data properties* dentro da classe *Biological Oceanography*, que, por sua vez, era uma subclasse de *Observation* (criada para abarcar descritores de dados observacionais), como mostram as Figuras 2 e 3. Para formalização da ontologia, foi utilizado o software *Protégé*<sup>5</sup>.

<sup>5</sup> Disponível em: <<http://protege.stanford.edu/>>. Acesso: 20 jul. 2017.

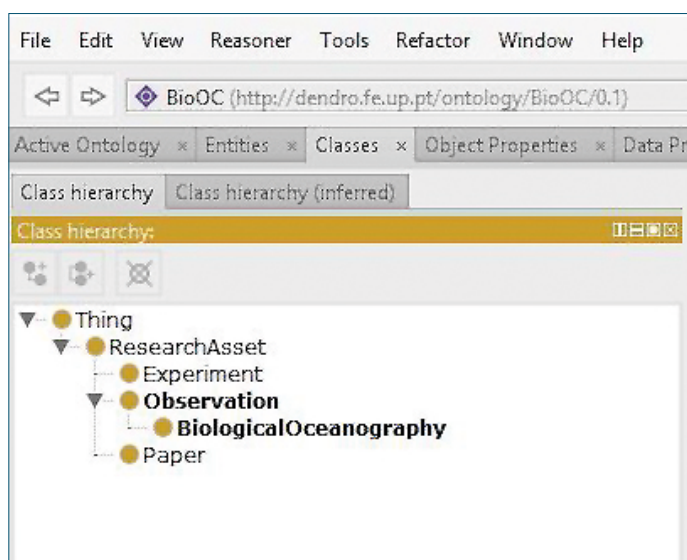


Figura 2 – Classe *Biological Oceanography/observation*  
 Fonte: Autoria própria (2015).

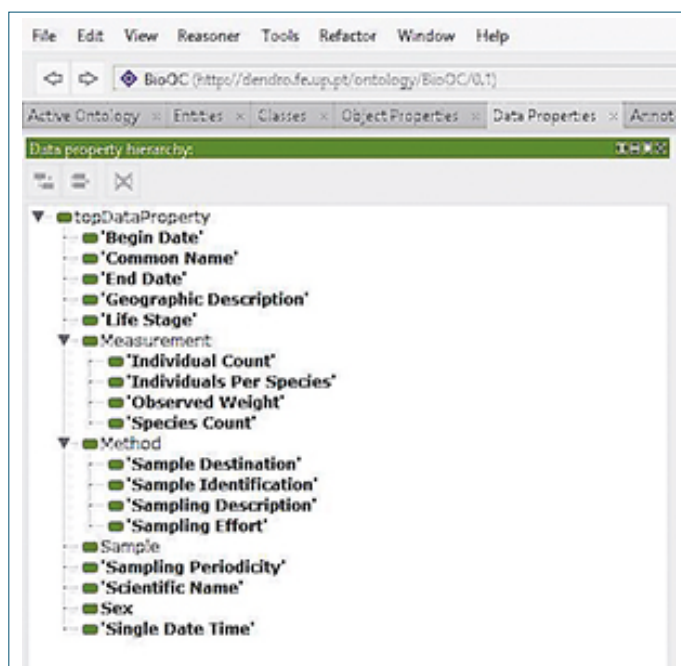


Figura 3 – Propriedades da ontologia *Biological Oceanography*  
 Fonte: Autoria própria (2015).



O vocabulário *Biological Oceanography* foi então carregado para a plataforma para compor, juntamente com vocabulários de outros domínios, um painel multidisciplinar de descritores para documentação de conjuntos de dados de investigação. Os investigadores, ao depositarem seus dados na plataforma Dendro, têm a possibilidade de escolher os descritores de qualquer um dos vocabulários ingestados na plataforma, formando seus próprios perfis de aplicação ao customizar a anotação de seus dados conforme sua necessidade. Obviamente, a tarefa de descrever dados torna-se mais simples para o investigador diante de um vocabulário que ele próprio ajudou a construir.

A Figura 4 apresenta um exemplo do vocabulário *Biological Oceanography* figurando na plataforma Dendro juntamente com os vocabulários de outros domínios.

The screenshot displays the Dendro platform's data entry interface. At the top, there's a navigation bar with 'Dendro®', 'My', 'Find', 'Plugins', 'Search Dendro', 'SUBMIT', and a user profile 'Rubia Gattelli'. Below this, a 'Selection' panel on the left shows a file upload area with 'Up to teste01' and a PDF file '24723-Codebook.pdf'. The main form area contains several fields: 'funded by' (Robert Wood Johnson Foundation (57814)), 'Specimen' (Sample: Over the life of the initiative, each grantee was expected to recruit 900 par...), 'Organization name' (Inter-university Consortium for Political and Social Research), 'Begin Date' (2003), 'End Date' (2007), 'Title' (Active for Life: Translation of Physical Activity Programs for Mid-Life and Older Adult), and 'Abstract'. On the right, a sidebar lists various vocabularies: 'Friend of a friend', 'Dendro research', 'Double Cantilever Beam', 'Pollutant analysis', 'Biodiversity evolution studies', 'Biological Oceanography', and 'Gravimetry'. Each vocabulary entry includes a brief description of its scope.

Figura 4 – Vocabulários multidomínio para descrição de dados na plataforma Dendro  
Fonte: Autoria própria (2015).



## CONSIDERAÇÕES FINAIS

A curadoria digital para dados de investigação pode ser considerada, no atual contexto científico, uma tarefa de suma importância. O depósito de conjuntos dos dados em plataformas seguras, bem como a adoção de metadados descritivos assegura que os dados estejam preservados, possam ser entendidos e reutilizados em futuros projetos de investigação. Essas ações trazem uma série de vantagens para o mundo acadêmico e científico, dentre elas, a transparência nas práticas e resultados de investigação, economia nas coletas de dados, colaboração entre cientistas e preservação histórica de dados.

O desenvolvimento de modelos de metadados específicos para domínios pode contribuir para que os investigadores invistam na documentação dos dados logo que gerados. A tarefa de documentar torna-se mais simples, à medida que os vocabulários são construídos com a colaboração dos investigadores, evitando barreira de linguagem e processos complexos de depósito. Isso demonstra que o trabalho conjunto entre curadores e investigadores é eficiente e traz resultados positivos para formar uma cultura de curadoria de dados de investigação.

Por fim, cabe ressaltar que pode ser largamente empregada a metodologia adotada para o desenho de modelos de metadados específicos de domínios, a formalização dos descritores em ontologias e sua adoção em plataforma de gestão de dados de modo que funcionem como perfis de aplicação, considerando as adaptações necessárias. Os profissionais da informação possuem as habilidades necessárias para atuar como curadores, não somente nas questões mais técnicas (como na elaboração de perfis de aplicação), mas também na disseminação das boas práticas de gestão de dados.

## REFERÊNCIAS

ABBOTT, D. **Annotation**. 2008. Disponível em: <<http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/annotation%20>>. Acesso em: 29 dez. 2014.

CASTRO, J. A. et al. Involving data creators in an ontology-based design process for metadata models. In: MALTA, M. C.; BAPTISTA, A. A.; WALK, P. (Ed.). **Developing metadata application profiles**. Hershey, Pensilvânia: IGI Global, 2016. p. 181-214.

CASTRO, J. A.; RIBEIRO, C.; SILVA, J. R. da. Designing an application profile using qualified Dublin Core: a case study with fracture mechanics datasets. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2013, Lisboa. **Proceedings...** Lisboa: *Dublin Core Metadata Initiative, 2013*. Disponível em: <<http://dcpapers.dublincore.org/pubs/article/view/3685>>. Acesso em: 30 set. 2014.

CASTRO, J. A.; SILVA, J. R. da; RIBEIRO, C. Creating lightweight ontologies for dataset description: practical applications in a cross-domain research data management workflow. In: ACM/IEEE -CS JOINT CONFERENCE ON DIGITAL LIBRARIES, 14th., 2014, London. **Proceedings...** Londres: IEEE Press, 2014. Disponível em: <<http://dl.acm.org/citation.cfm?id=2740823>>. Acesso em: 25 maio 2017.

CORTI, L. et al. **Managing and sharing research data: a guide to good practice**. Los Angeles: SAGE, 2014.

DUVAL, E. et al. Metadata principles and practicalities. **D-Lib Magazine**, v. 8, n. 4, april, 2002. Disponível em: <<http://www.dlib.org/dlib/april02/weibel/04weibel.html>>. Acesso em: 10 mar. 2017.

GORE, S. eScience and data management resources on the web. **Medical Reference Services Quarterly**, v. 30, n. 2, p. 167-177, 2011.

HEERY, R.; PATEL, M. Application profiles: mixing and matching metadata schemas. **Ariadne**, v. 25, sept. 2000. Disponível em: <<http://www.ariadne.ac.uk/issue25/app-profiles>>. Acesso em: 22 out. 2014.

OSSWALD, A. E-science and information services: a missing link in the context of digital libraries. **Online Information Review**, United Kingdom, v. 32, n. 4, p. 516-523, 2008. Disponível em: <<http://www.emeraldinsight.com/doi/abs/10.1108/14684520810897395>>. Acesso em: 24 mar. 2017.

PENNOCK, M. **Curating e-Science data**. Edinburgh: Digital Curation Centre, 2006. Disponível em: <<http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/curating-e-science-data>>. Acesso em: 29 dez. 2014.

QIN, J.; D'IGNAZIO, J. The central role of metadata in a science data literacy course. **Journal of Library Metadata**, v. 10, n. 2/3, p.188-204, ago. 2010. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/19386389.2010.506379>>. Acesso em: 24 mar. 2017.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: Comissão Nacional de Energia Nuclear, 2015. Disponível em: <[http://carpedien.ien.gov.br/bitstream/ien/1624/1/GUIA\\_DE\\_DADOS\\_DE\\_PESQUISA.pdf](http://carpedien.ien.gov.br/bitstream/ien/1624/1/GUIA_DE_DADOS_DE_PESQUISA.pdf)>. Acesso em: 04 jun. 2016.

WHITE, H. C. Descriptive metadata for scientific data repositories: a comparison of information scientist and scientist organizing behaviors. **Journal of Library Metadata**, v. 14, n. 1, p. 24-51, jan. 2014. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/19386389.2014.891896>>. Acesso em: 24 mar. 2017.

WHITMIRE, A. Thoughts on "eResearch": a scientist's perspective. **Journal of eScience Librarianship**, v. 2, n. 2, p. 68-72, 2013. Disponível em: <<http://escholarship.umassmed.edu/jeslib/vol2/iss2/3/>>. Acesso em: 09 dez. 2014.

# GESTÃO DE DADOS MULTIDISCIPLINAR NA UNIVERSIDADE DO PORTO: EXPERIÊNCIAS EM GRUPOS DE DIMENSÕES DIVERSAS

Cristina Ribeiro  
João Rocha da Silva  
João Aguiar Castro  
Ricardo Carvalho Amorim  
João Correia Lopes  
Gabriel David





## INTRODUÇÃO

Os dados de investigação são criados e usados em diversos contextos. Podem ser gerados especificamente para um projeto, como no caso de dados de sensores captados em trabalho experimental, ou de dados de entrevistas para uma pesquisa. Incluem dados que são captados sistematicamente para algum fim e que também são usados em investigação, como é o caso dos dados meteorológicos ou dos registos de segurança de um centro de computação. Podem ser dados banais, como um conjunto de páginas web, colecionadas *ad-hoc* para estudar o desempenho de uma ferramenta de busca automática. Esta diversidade torna a gestão de dados de investigação (GDI) uma tarefa difusa, para a qual os investigadores não têm processos bem estabelecidos nem uma intuição clara quanto à sua utilidade.

Os dados gerados em grandes projetos com áreas de investigação bem financiada já estão a ser tratados em infraestruturas disciplinares. O *National Center for Biotechnology Information* (NCBI) (NCBI Resource Coordinators, 2013), nas ciências da vida, e o *Inter-university Consortium for Political and Social Research* ICPSR (DOTY et al., 2015), nas ciências sociais, são bons exemplos de infraestruturas estáveis que suportam comunidades grandes e que são usadas por investigadores tanto para obter dados de base como para contribuir com novos resultados.

Para os dados na chamada **cauda longa** da ciência, não há de momento soluções genéricas para a GDI e, mesmo quando os investigadores reconhecem o valor da gestão de dados, é habitual não terem apoio para os tornar visíveis, satisfatoriamente descritos, depositados e pesquisáveis. Os esforços nesta área estão de momento em estado de projeto, tal como ilustrado por duas grandes iniciativas de financiamento europeu, o *Open Access Infrastructure for Research in Europe* (OpenAIRE) (MANGHI et al., 2010) e o *European Data Infrastructure* (EU-DAT) (LECARPENTIER; MICHELINI; WITTENBURG, 2013).

Um aspeto importante a ter em conta são os atores envolvidos na GDI e as condições que fomentam a sua colaboração (RIBEIRO et al., 2015). A Figura 1 mostra esses atores e as ferramentas que usamos nos vários passos do processo de investigação. Colocamos o foco nos papéis dos investigadores, gestores de investigação e curadores, e na sua colaboração com desenvolvedores para construir um suporte eficaz para a descrição e a publicação de dados.

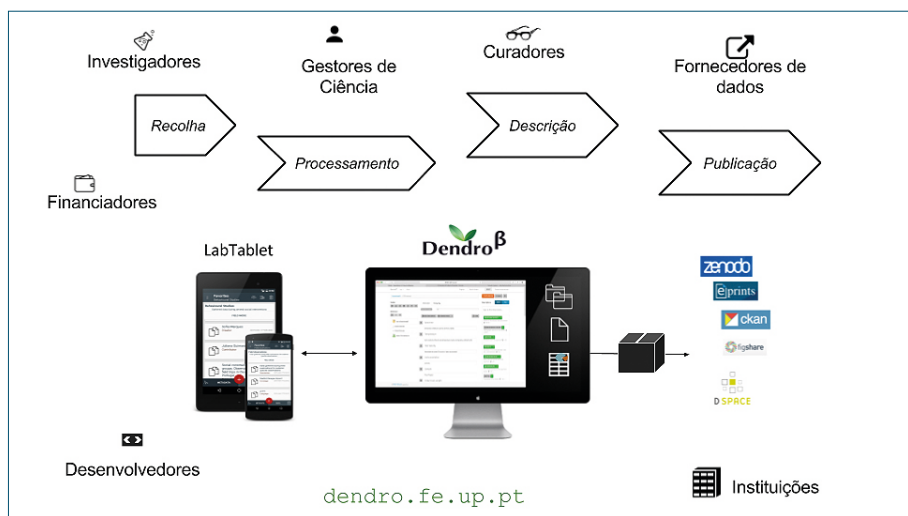


Figura 1 – GDI Atores, passos do fluxo de gestão de dados e ferramentas

Fonte: Autoria própria (2016).

O trabalho em GDI na Universidade do Porto começou com um estudo de avaliação, com a colaboração de 8 grupos de investigação, usando as recomendações existentes a nível internacional e cobrindo aspetos como a sensibilidade relativamente à gestão de dados, as necessidades mais urgentes relativamente aos conjuntos de dados existentes, as soluções existentes para armazenamento de dados, o valor estimado para dados legados, e as necessidades de apoio para ações de GDI (RIBEIRO; FERNANDES 2011). Com base neste estudo, começámos a desenhar um fluxo de trabalho para grupos de investigação e as ferramentas para melhorar a eficácia da publicação e disseminação de investigação.

A preocupação com a GDI na cauda longa é relativamente recente, e diversas entidades fornecem já suporte para dados multi-domínio (*Digital Curation Centre*<sup>1</sup>, *Australian National Data Service*<sup>2</sup>, *Data Archiving and Networked Services*<sup>3</sup>, *Data Observation Network for Earth*<sup>4</sup>, *Dash*<sup>5</sup>). No entanto, em algumas áreas em que os grandes conjuntos de dados são essenciais, há um historial mais longo de iniciativas GDI, como sejam a organização de bases de dados que identificam contribuições individuais, com uma ligação estabelecida às publicações. O NCBI e o ICPSR foram já mencionados e são casos destes nas ciências da vida e nas ciências sociais, respetivamente. Noutras áreas, à medida que os grupos internacionais crescem, a partilha de dados se torna mais impor-

1 Disponível em: <<http://www.dcc.ac.uk/>>. Acesso em: 22 jul. 2017.

2 Disponível em: <<http://ands.org.au/>>. Acesso em: 22 jul. 2017.

3 Disponível em: <<http://www.dans.knaw.nl/en>>. Acesso em: 22 jul. 2017.

4 Disponível em: <<https://www.dataone.org/>>. Acesso em: 22 jul. 2017.

5 Disponível em: <<https://dash.cdlib.org/>>. Acesso em: 22 jul. 2017.

tante, e a avaliação das contribuições requer a ligação ao material de base, novas infraestruturas estão a ser estabelecidas. Isto é ilustrado pelos projetos promovidos pelo *European Strategy Forum on Research Infrastructures* (ESFRI), que apoia uma rede de Infraestruturas de Investigação de interesse pan-europeu (EUROPEAN STRATEGY FORUM ON RESEARCH INFRASTRUCTURES, 2016).

As ferramentas disponíveis para os investigadores são decisivas para o seu envolvimento. Para grupos na cauda longa, o nosso foco é na preparação dos dados, nomeadamente a organização dos conjuntos de dados e a criação de metadados. O fluxo de trabalho proposto neste caso inclui o Dendro, uma plataforma de gestão de dados baseada em ferramentas de código aberto, e o LabTablet, um caderno de laboratório eletrónico para automatizar a criação de metadados. Para grupos que precisam de abordagens específicas, a nossa análise resultou no desenvolvimento de modelos e aplicações para organizar os dados e apoiar os seus casos de uso principais. Em todos os casos, adotámos as ontologias para modelação dos metadados, tendo em vista a disseminação de metadados como dados abertos ligados (*linked open data*).

## GRUPOS DE INVESTIGAÇÃO E REQUISITOS DE GESTÃO DE DADOS DE INVESTIGAÇÃO

A necessidade de acesso e reutilização de dados é reconhecida, e o fato de haver pouca reutilização pode relacionar-se com a dificuldade sentida pelos investigadores em criarem suficientes metadados contextuais (FANIEL; YAKEL, 2011). A razão para este fato pode ser que os investigadores estão em geral orientados para produzir documentação para os dados de acordo com as suas próprias necessidades (MAYERNIK, 2011).

Relativamente à documentação, é possível seguir uma abordagem tradicional, descrevendo os dados de acordo com modelos já estabelecidos para as publicações, e depositando-os em repositórios existentes, como mais um tipo de publicação. Esta é a linha seguida em iniciativas como o projeto europeu OpenAIRE<sup>6</sup>, promovendo o repositório Zenodo<sup>7</sup> para recolher e interligar artigos, conjuntos de dados, software e os projetos que lhes dão origem. Esta abordagem torna a tarefa de descrição mais leve, mas a falta de descrição de dados dependente do domínio pode torná-los mais difíceis de localizar, interpretar e reutilizar. Um investigador interessado num conjunto de dados, mesmo que o localize, terá provavelmente de contactar os autores para compreender o seu conteúdo.

6 Disponível em: <<https://www.openaire.eu/>>. Acesso em: 22 jul. 2017.

7 Disponível em: <<https://zenodo.org/>>. Acesso em: 22 jul. 2017.



Como alternativa, propomos uma abordagem em que os investigadores são munidos de modelos de metadados expressivos para descrever os seus dados. Isto requer ferramentas, mas também o desenho de processos que requerem algum apoio especializado. Nesta proposta definimos um modelo de metadados distribuído e multi-domínio como parte do fluxo de curadoria, o que requer a identificação de requisitos de GDI em cada domínio. Para este efeito identificam-se os conceitos do domínio e representam-se como descritores que podem ser combinados na interface do Dendro para os investigadores descreverem os seus dados com informação contextual e completa (SILVA; RIBEIRO; LOPES, 2016).

Para atingir este objetivo tivemos a colaboração de um conjunto de grupos de investigação de diversas áreas dentro dos quais pedimos aos investigadores algum empenhamento em atividades de GDI. Os contatos preliminares foram parte de um estudo de análise (SILVA; RIBEIRO; LOPES, 2011) que seguiu as recomendações do Data Asset Framework (2016), para identificar a natureza e objetivos dos conjuntos de dados existentes e apresentar aos investigadores os conceitos da GDI. O método envolveu a identificação de conjuntos de dados relevantes para publicação e a avaliação de ferramentas de GDI no contexto das atividades regulares dos grupos.

Para promover a descrição de dados tão cedo quanto possível no fluxo de investigação, procuramos envolver os investigadores no desenvolvimento de modelos de metadados específicos do domínio. As áreas de trabalho dos grupos de investigação vão desde configurações de teste que produzem dados experimentais, como é o caso das experiências de *double cantilever beam*, a cenários mais intensivos em computação como a simulação do desempenho de veículos (CASTRO; SILVA; RIBEIRO, 2014; CASTRO et al., 2015), os problemas de otimização combinatória em investigação operacional ou os estudos de ecologia preditiva no domínio da biodiversidade. A ligação a estes grupos usou diversas técnicas, nomeadamente entrevistas, análise de conteúdos em publicações e experiências de descrição de dados. As reuniões em formato livre foram usadas para recolher tanto quanto possível a visão dos investigadores.

As entrevistas semi-estruturadas são habitualmente seguidas nas primeiras interações com os investigadores, permitindo obter conhecimento sobre o domínio e as práticas de GDI no ciclo de vida dos dados no grupo respectivo. O guião de entrevista foi adaptado do *Data Curation Profile Toolkit* e traduzido para o português. Nos primeiros casos de estudo (mecânica de fratura, química analítica) as entrevistas foram realizadas com pouca informação prévia aos entrevistados. Nos casos seguintes o formulário de entrevista foi enviado previamente por correio eletrónico, para que os investigadores tivessem oportunidade de o ler e de preparar respostas, garantindo informação mais completa.



A análise de conteúdos nas publicações dos investigadores serve de complemento à entrevista, uma vez que as publicações são boas fontes de informação sobre a recolha e análise de dados. As seções de metodologia, por exemplo, habitualmente articulam informação que pode contextualizar os dados, como configurações experimentais, características do ambiente e variáveis usadas. Isto foi bastante interessante, por exemplo, no domínio da simulação de veículos (CASTRO et al., 2015). Nalguns casos, como o da produção de hidrogénio, a análise de conteúdo foi feita antes da primeira reunião com os investigadores. No domínio da dinâmica de fluidos computacional, a análise de conteúdo foi a principal fonte de identificação de conceitos, por falta de disponibilidade dos investigadores. Neste caso os especialistas do domínio foram consultados para validar os conceitos selecionados logo na primeira reunião.

Todos os grupos de investigação participaram numa experiência de descrição de dados no Dendro, em que foram convidados a criar pastas, carregar ficheiros e escolher descritores dos vários modelos de metadados disponíveis. Nesta experiência participaram pelo menos 2 investigadores de cada grupo, um usando uma versão do Dendro com todos os descritores organizados por vocabulários, o outro tendo disponível uma ordenação de descritores de acordo com o uso que estes tiveram anteriormente (SILVA; RIBEIRO; LOPES, 2016).

No que diz respeito a reuniões em formato livre, estas tiveram lugar tanto antes como depois das entrevistas semi-estruturadas, dependendo da disponibilidade dos investigadores e da facilidade para reunião. Agendámos reuniões preliminares para apresentar os objetivos das interações subsequentes e para recolher os requisitos de GDI e as perspetivas dos investigadores. Com os grupos em que não foi possível uma análise mais aprofundada, aproveitámos para algum contato informal após a própria entrevista. Estas reuniões foram úteis para aprofundar os pontos principais da entrevista ou para identificar a informação requerida para descrever os conjuntos de dados que estavam a ser trabalhados.

Relativamente aos metadados, enquanto a maior parte dos grupos não estava familiarizada com o conceito, o grupo de biodiversidade já estava a trabalhar em recomendações de metadados da diretiva europeia *Infrastructure for Spatial Information in Europe* (INSPIRE)<sup>8</sup> no contexto de um projeto em curso. Neste caso pudemos mapear os conceitos INSPIRE para desenvolver o modelo de metadados para o domínio. O mesmo se passou no domínio das ciências sociais, onde foi possível reutilizar conceitos da *Data Documentation Initiative* (DDI)<sup>9</sup>. Os descritores para a oceanografia foram escolhidos nos vocabulários da *Ecological Metadata Language*<sup>10</sup> e *Darwin Core*<sup>11</sup>. Para outros domínios não há vocabulários

8 Disponível em: <<http://inspire.ec.europa.eu/>>. Acesso em: 22 jul. 2017.

9 Disponível em: <<https://www.ddialliance.org/>>. Acesso em: 22 jul. 2017.

10 Disponível em: <<https://knbn.ecoinformatics.org/#external/emlparser/docs/index.html>>. Acesso em: 22 jul. 2017.

11 Disponível em: <<http://rs.tdwg.org/dwc/>>. Acesso em: 22 jul. 2017.

já disponíveis, pelo que se iniciou a sua definição em colaboração com os investigadores.

Relativamente ao uso de ferramentas (Dendro e LabTablet, referidos a seguir), todos os grupos realizaram tarefas de descrição usando o Dendro e nalguns casos foi também usado o LabTablet, dependendo do contexto dos dados. Em grupos que recolhem dados em trabalho de campo ou que realizam experiências em ambientes controlados, o uso do LabTablet é mais natural do que nos que realizam estudos computacionais. O LabTablet foi experimentado por investigadores das ciências sociais e da biodiversidade.

O tempo dedicado ao trabalho com cada grupo é também relevante, até porque não temos um controlo estrito sobre este. A agenda da colaboração é difícil de prever, e a nossa experiência mostrou que a mesma quantidade de interação ou nível de envolvimento podem ser obtidos em períodos de tempo muito diferentes, dependendo do grupo. Como um exemplo, num dos casos realizámos três reuniões num período de três meses; as mesmas três reuniões, realizaram-se em apenas duas semanas num outro caso, com resultados semelhantes. Nalguns casos os grupos contactados tiveram apenas uma interação esporádica ou mais informal. Consideramos estes casos porque ajudam a compor um conjunto mais representativo dos domínios de investigação.

A avaliação preliminar do sucesso deste modelo de metadados distribuído e multi-domínio é apenas baseada em comentário informal e nos resultados das tarefas de descrição.

## **ORGANIZAÇÃO DE DADOS E CRIAÇÃO DE METADADOS COM DENDRO E LABTABLET**

Os avanços na GDI dependem da existência de processos claros para os investigadores no que diz respeito a recolha de dados, organização, descrição e publicação ou depósito. As entrevistas com os investigadores acerca das suas práticas de GDI mostraram uma lacuna grande entre os processos usados na preparação de artigos e as rotinas requeridas para organizar dados e torná-los aptos para publicação. Ficou claro que as ferramentas são essenciais para facilitar a tarefa dos utilizadores em grupos pequenos, e começamos a desenhar, implementar e testar ferramentas para preparar dados na **cauda longa**. Estas ferramentas podem ter um campo de aplicação muito grande – recolha de dados, limpeza de dados, processamento, organização, descrição, depósito e pesquisa são exemplos. Como a GDI está a evoluir como parte integrante da investigação, é de esperar que as plataformas de repositórios, disciplinares ou genéricas, se tornem vulgares. As ferramentas para recolha e processamento de dados serão

possivelmente muito dependentes do domínio. Focamos a nossa atenção na organização e descrição de dados, que se situam na interface entre os investigadores e os gestores de repositórios. As ferramentas para preparar e descrever conjuntos de dados são importantes para estabelecer um discurso de GDI consistente com os investigadores, e as mesmas ferramentas permitem-nos fazer a interface com qualquer solução de repositório. Uma descrição rica é também muito importante para a pesquisa, uma vez que na maior parte dos casos os conjuntos de dados têm pouco ou nenhum conteúdo textual indexável.

O Dendro é uma plataforma GDI baseada em ontologias usada na Universidade do Porto, baseada em componentes de código aberto. Tem como alvo os investigadores, nas tarefas de depósito e partilha de dados tanto dentro dos seus grupos de investigação como com elementos externos. O Dendro usa a metáfora da Dropbox para carregamento de dados, juntando-lhe funcionalidades sofisticadas de descrição de dados. Os conceitos no Dendro incluem o Projeto – uma pasta partilhada onde os membros depositam ficheiros de vários tipos, a Pasta – os elementos do projeto criam pastas e subpastas para organizar dados, e o Descritor – os metadados são associados a recursos (pastas e ficheiros) usando descritores genéricos ou específicos do domínio. O Dendro integra-se no fluxo de investigação cobrindo os passos entre a criação dos dados e a sua publicação, e pode exportar dados e metadados, ou só os metadados, para as plataformas de repositório mais divulgadas, onde o grupo pode depois publicá-las se e quando assim o entender.

Os dispositivos móveis têm evoluído para incluir capacidades avançadas que os tornam adequados para um grande leque de atividades relacionadas com a investigação. Numerosos casos de investigadores que usam os seus dispositivos para melhorar o processo de investigação mostram que esta é uma tendência forte. Para além da crescente capacidade de armazenamento, os dispositivos móveis estão frequentemente ligados à Internet em permanência e têm sensores integrados que fornecem dados contextuais sobre o ambiente do investigador. Esta foi a motivação para o desenvolvimento do LabTablet, uma aplicação para dispositivos Android com ênfase na GDI.

O LabTablet segue a linha dos chamados cadernos de laboratório eletrônicos, aplicações que tiram partido dos sensores nos dispositivos móveis para auxiliar os investigadores a descrever os seus dados. Nalguns casos a descrição com o dispositivo móvel substitui um processo que antes usaria o caderno de laboratório, garantindo que os metadados não são perdidos mas sim registados, associados a dados e depositados. Uma ferramenta simples como esta, integrada em dispositivos de uso frequente, pode contribuir para gerar mais metadados associados a um conjunto de dados, com pouco esforço adicional. Além disso,

com o dispositivo móvel a criação de metadados torna-se uma parte indivisível da recolha de dados, distribuindo o esforço de descrição ao longo do projeto e evitando tarefas consumidoras de tempo no final do projeto. O fluxo de organização e descrição de dados com o Dendro e o LabTablet está ilustrado na Figura 2, com um projeto no domínio da biodiversidade. Podemos ver aqui a estrutura de pastas e ficheiros do projeto, os descritores usados neste domínio (*Dublin Core* genérico mais um subconjunto de descritores INSPIRE selecionados para este domínio) e a comunicação entre o Dendro e o LabTablet: modelos de metadados são passados do Dendro para o LabTablet, a recolha de metadados no LabTablet preenche esses descritores, e os valores obtidos são carregados no Dendro quando as plataforma sincronizam.

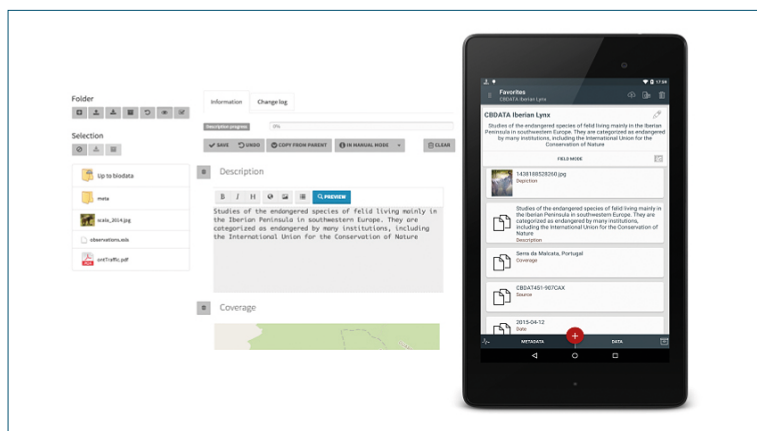


Figura 2–Organização e descrição de dados com o Dendro e o LabTablet  
Fonte: Autoria própria (2016).

A eficácia da plataforma Dendro foi testada em diversas experiências. Três destas realizaram-se em diferentes condições, e os resultados estão disponíveis em dois dos casos. A primeira experiência, DendroDC1, usou uma versão preliminar do Dendro, depois de ter sido testada com alguns investigadores do nosso painel. Os participantes foram estudantes de Mestrado em Ciência da Informação, a quem se pediu que preenchessem alguns descritores *Dublin Core* usando dados provenientes de diversas fontes. A experiência teve dois objetivos: testar o Dendro em condições realistas de carga e observar o uso de um modelo genérico de metadados na anotação de dados usando recomendação de descritores com base no uso. O sistema provou ser robusto e mostrou melhorias na descrição com o uso de recomendação (SILVA; RIBEIRO; LOPES, 2016). A segunda experiência, DendroPanel, foi mais realista e envolveu 11 grupos e instâncias de Dendro

configuradas com ontologias para todos os seus domínios, quando disponíveis. Para cada grupo tivemos dois participantes, que descreveram conjuntos de dados do seu domínio: um deles usou o Dendro com recomendação, e o outro sem. Os resultados foram favoráveis ao uso da ordenação de descritores, que funcionou como uma recomendação baseada no uso. A terceira experiência, DendroDC2, repetiu a primeira com um novo grupo de participantes.

O LabTablet também foi testado com alguns investigadores em engenharia e na biodiversidade. Esta plataforma requer ainda experiências mais alargadas. Entretanto, a plataforma já foi usada como base para duas aplicações relacionadas com dados: uma para mostrar as condições do mar para a prática de desportos náuticos (AMORIM et al., 2016) e outra para apoiar a recolha de dados de espécimes no projeto *SeaBioData*<sup>12</sup>, mencionado a seguir.

## **SOLUÇÕES DISCIPLINARES: MODELOS DE DADOS E DE METADADOS DEDICADOS**

À medida que explorámos os requisitos de grupos de investigação numa instituição de investigação muito grande, com o foco em grupos pequenos em que a gestão de dados ainda tem de ser estabelecida como um processo ágil, fomos confrontados com casos que não encaixam neste padrão.

As soluções para estas áreas levaram a projetos específicos que incluíram a análise, desenho e implementação de soluções que respondem a requisitos mais específicos. Vamos referir a seguir três desses projetos que se desenvolveram como iniciativas individuais, envolvendo equipas das disciplinas correspondentes, e tendo a GDI no centro. Estes projetos dão-nos uma visão sobre as funções desempenhadas por um sistema específico para um domínio, alargando o foco da organização e descrição para funções tais como a recolha de dados, o processamento e o arquivo.

## **DADOS DE SENSORES EM ESTRUTURAS DE BETÃO**

O método de investigação na área da monitorização da condição de estruturas de engenharia civil está bem estabelecido. Para monitorizar uma ponte, uma torre ou outra grande estrutura, o sistema de recolha de dados é desenhado e montado na estrutura e as sequências de dados com períodos da ordem dos milissegundos começam a ser produzidas, mantendo-se a recolha por meses ou anos. Cada amostra na sequência tem valores obtidos dos sensores do sistema de aquisição (temperatura, aceleração, entre outros). Cada fluxo é segmentado

<sup>12</sup> Disponível em: <<http://eeagrants.org/project-portal/project/PT02-0017>>. Acesso em: 22 jul. 2017.

em ficheiros para efeitos de transferência e armazenamento. Estes ficheiros crus são limpos, e o resultado é sujeito a uma ou mais rotinas de processamento. Os dados limpos e os resultados são visualizados, e as conclusões sobre o estado das estruturas são incorporadas em relatórios e em artigos científicos. No passado era comum, depois de os resultados serem publicados, os dados respectivos serem descartados.

A crescente sensibilidade para o valor dos dados, que constituem evidência sobre as conclusões publicadas e que servem como base para novos estudos, motivou a equipe a lançar um projeto de arquivo digital à medida das suas necessidades. Os requisitos foram levantados pela nossa equipe com um grande envolvimento da equipe disciplinar, resultando numa arquitetura do arquivo digital com três componentes. Uma estrutura sistemática de pastas com regras de nomeação organiza pastas e ficheiros no sistema de ficheiros; a base de dados recolhe todos os metadados; uma aplicação junta funcionalidades de visualização e de gestão.

Os metadados estão organizados em cinco pacotes:

- a) metadados contextuais sobre as características da estrutura e os pormenores do projeto, datas e fonte de financiamento;
- b) informação de sistema sobre o desenho, implementação e parametrização do sistema de aquisição de dados, os seus sensores e componentes;
- c) estrutura dos ficheiros de dados, variáveis recolhidas e suas unidades, frequência de amostragem, carimbo temporal e localização de ficheiros;
- d) atores envolvidos, tais como detentor e projetista da estrutura, equipe de investigação e investigadores externos;
- e) documentos de qualquer natureza, como: contrato do projeto, desenho do sistema de aquisição, fotos, relatórios e artigos publicados.

A aplicação é baseada na web e tem funcionalidades como a gestão dos cinco pacotes de metadados, a ingestão automática de novos ficheiros de dados, uma ferramenta de visualização gráfica para navegar os dados por intervalos e uma função assíncrona para exportar ficheiros de dados usando *Open Archives Initiative – Protocol for Metadata Harvesting* (OAI-PMH). O sistema desenvolvido (ViBest SHM<sup>13</sup>) é suficientemente genérico para monitorar sistemas em outros domínios para além das condições de saúde de estruturas na engenharia civil.

---

13 Disponível em: <<http://vibest.fe.up.pt/shm>>. Acesso em: 22 jul. 2017.

## DADOS DE VENTO OBTIDOS DE LIDARS

O *Winds@UP* é um protótipo de e-infraestrutura desenvolvido no âmbito do projeto *WindScanner.eu* e inclui um repositório para conjuntos de dados experimentais. Estes são séries de dados georreferenciados gerados pelos sensores LIDAR usados em testes de campo ao ar livre (campanhas de campo). A plataforma gera os dados e metadados correspondentes, fornece capacidades de processamento para processamento de dados in situ e está instalada no *WindScanner Hub* (GOMES et al., 2014).

O dispositivo *WindScanner* é um LIDAR de vento (de curto e longo alcance) que é usado em conjuntos coordenados de 3 unidades para medir campos vectoriais de vento 3D com alta precisão. É a tecnologia central da infraestrutura *WindScanner* a ser usada no projeto ESFRI correspondente. Os *WindScanners* são instalados em locais de teste existentes ou planeados, cobrindo diferentes condições climáticas e terrenos. Há procedimentos normalizados que podem ser aplicados aos dados recolhidos, e as séries temporais resultantes são armazenadas para uso futuro. Todas as séries temporais têm marcas *Global Positioning System* (GPS) usadas para alinhar os sinais obtidos de dispositivos diferentes. Além dos dados base e dos derivados, a plataforma *Winds@UP* fornece armazenamento para metadados e outros recursos – objetos de investigação – criados pelos investigadores no decorrer do seu trabalho. Os recursos incluem conjuntos de dados, arquivos, fotos, gráficos e artigos científicos.

Os metadados dos objetos de investigação são organizados em três categorias:

- a) descritivos: título, autor, resumo e palavras-chave, que ajudam a descoberta de recursos através de pesquisa e navegação;
- b) administrativos: preservação, gestão de direitos, aspetos técnicos como formatos ou configuração experimental;
- c) estruturais: como os diferentes componentes de um conjunto de objetos de dados associados se relacionam entre si (conjuntos de dados, procedimentos e resultados).

Os dados e metadados associados, empacotados em formatos normalizados e independentes das máquinas, são carregados no repositório da plataforma. Neste ponto, os objetos de investigação são descritos usando o mesmo modelo de metadados (representado como uma ontologia) usado para outros objetos de investigação do domínio. Dados brutos, limpos e processados são usados, em conjunto com objetos de investigação relacionados, por um portal web com facilidades de pesquisa apropriadas para os investigadores, facilitando a colaboração e a reprodução de resultados.

O WindS@UP foi desenhado como uma e-infraestrutura que fornece armazenamento dedicado e local a uma comunidade. Isto distingue-o dos casos **cauda longa**, que usam repositórios externos para armazenamento. Os conceitos de campanha, experiência, fenómeno observado e série temporal são comuns a outros domínios, tais como os elementos de metadados genéricos usados para os representar.

O *WindS@UP* é neste momento uma plataforma sólida, que passou por três ciclos de desenvolvimento e que contribuiu para a fase de preparação de uma infraestrutura de investigação ESFRI. Espera-se continuar o seu desenvolvimento no contexto da próxima fase da estrutura europeia ESFRI de investigação do vento. A versão mais recente do protótipo (*windsP*) foi usada recentemente para planear as experiências do projeto europeu *New European Wind Atlas* (NEWA)<sup>14</sup> e acompanhará a execução da campanha de campo em Perdigão, uma experiência de dupla colina em Portugal.

## DADOS DO MAR E DA ATMOSFERA

A investigação marinha é tipicamente organizada em campanhas a bordo de navios de investigação. Uma campanha segue um conjunto de estações em locais predefinidos. Em cada estação são recolhidas amostras do solo e da coluna de água usando dispositivos adequados. As amostras podem ser alvo de exame no local e, mais tarde, ser complementadas com análises físicas, químicas e biológicas, seguindo metodologias apropriadas. Os dados recolhidos são de natureza muito diversa, desde variáveis medidas em unidades especificadas, a identificação da presença ou densidade de espécies biológicas, imagens, vídeo e som, até espécimes capturados. A georreferenciação e o registro detalhado do processo de recolha das amostras e dos investigadores envolvidos são muito importantes. Ao mesmo tempo, há sequências de dados recebidas de instrumentos instalados em boias, de posições de navios e de dados de satélite que ficam fora do cenário das campanhas.

O projeto *SeaBioData*<sup>15</sup> surgiu dos contatos com uma equipe grande de investigação marinha e resultou na construção de um arquivo digital de dados cujo modelo contempla os conceitos relevantes já referidos.

A área da investigação marinha, como parte da investigação em ambiente, é fortemente regulada por regras europeias e nacionais como a diretiva INSPIRE (EUROPEAN COMMISSION JOINT RESEARCH, 2013) e as que emanam do Sistema Nacional de Informação do Mar (SNIMar)<sup>16</sup>. A diretiva europeia INSPIRE

14 Disponível em: <<http://www.neweuropeanwindatlas.eu/>>. Acesso em: 22 jul. 2017.

15 Disponível em: <<http://eeagrants.org/project-portal/project/PT02-0017>>. Acesso em: 22 jul. 2017.

16 Disponível em: <<http://www.snimar.pt/ar/ficheiros/perfilSNIMar.pdf>>. Acesso em: 22 jul. 2017.



impõe uma interface de programação para aplicações (API) muito estrita aos repositórios da área do ambiente para favorecer a interoperabilidade. Esta API está baseada num modelo de dados abstrato centrado em observações com valores para propriedades, métodos associados e referências geográficas. Das implementações deste modelo de dados, escolhemos duas: a implementação *52North Sensor Observation Service* (SOS) e a implementação *Geoserver Web Feature Service* (WFS), *Web Map Service* (WMS), e *Web Coverage Service* (WCS).

Este projeto resultou na criação de um cliente para estes dois servidores que alinha com os conceitos, folhas de recolha de dados e métodos da equipe de investigação marinha e que tem capacidade para adquirir os dados já existentes em formato digital e para os apresentar de forma visualmente efetiva.

## OS REQUISITOS DAS INFRAESTRUTURAS DISCIPLINARES

Em plataformas disciplinares, como as do *WindScanner.eu* e do *SeaBioData*, para além dos requisitos que já estamos a satisfazer, é provável que, à medida que mais grupos de investigação participem e usem a plataforma, sejam necessárias mais funcionalidades. Isso pode ser visto como uma evolução natural das infraestruturas e mesmo como um sinal do sucesso na sua adoção.

Um efeito que pode resultar da definição das infraestruturas disciplinares decorre da sua especialização: os conjuntos de dados são geridos em estruturas específicas que, mesmo sendo abertas, se podem tornar difíceis de descobrir e de pesquisar. Dois tipos de desenvolvimentos atuais contrariam este efeito. Um é a existência de agregadores de repositórios de dados, tais como o *re3data*, apoiado pelo DataCite (BRASE; SENS; LAUTENSCHLAGER, 2015). O outro é a tendência da aplicação dos princípios dos dados ligados (*linked open data*, LOD) aos metadados. Se os metadados podem ser recolhidos por serviços LOD, aumenta a facilidade com que os dados se tornam acessíveis a diferentes tipos de aplicações.

## CONSIDERAÇÕES FINAIS

Quando olhamos para a GDI em todas as suas vertentes, temos de reconhecer que se trata de uma tarefa esmagadora. De certa forma, isso é comum a todas as iniciativas com uma componente de arquivo: como podemos estimar quais recursos vão ser valiosos, se nem todos podem ser preservados em longo prazo? Relativamente aos arquivos tradicionais, as iniciativas de arquivo na web já lidaram com o problema de tornar as tarefas de seleção e descrição mais leves e semiautomáticas. Os arquivos de dados de investigação têm um problema se-

melhante, no sentido de que algumas coleções serão muito mais preparadas que outras. No estado corrente das iniciativas em GDI, precisamos atingir um equilíbrio entre o esforço requerido para a descrição de dados, que torna os conjuntos de dados parte dos resultados de investigação, publicáveis e reutilizáveis, e as compensações imediatas que os investigadores recebem por investir na curadoria dos dados. No trabalho apresentado, defendemos as ferramentas que aligeiram o trabalho dos investigadores, embebendo a curadoria de dados no conjunto das atividades de investigação e fornecendo confiança nos resultados de investigação obtidos. No entanto, as infraestruturas aparecem a um ritmo mais rápido do que os investigadores podem acompanhar, e o fosso entre as suas funcionalidades e os requisitos apercebidos pelos investigadores é grande. A segunda linha de ação é, por isso, a análise de requisitos para a GDI, o seu alinhamento com as ferramentas existentes e a identificação dos componentes em falta.

No trabalho apresentado, há um foco em soluções para a **cauda longa**, mas também alguns resultados preliminares relativamente à GDI em áreas que requerem soluções à medida. É importante admitir desde o início que nem todos os grupos de investigação precisam do mesmo tipo de soluções, mas ainda assim procurar o que é comum a todos. Vimos os exemplos de um grupo de engenharia civil que coleciona dados de sensores, de uma comunidade de investigação do vento a preparar a infraestrutura para suportar a recolha e processamento de conjuntos de dados grandes, e de um instituto do mar e da atmosfera que lida com a diversidade de dados usados em produtos de previsão do clima, investigação marinha e recolha de espécimes. Para cada um destes grupos, desenhamos e implementámos sistemas dedicados que lidam com parte dos seus fluxos de trabalho. Uma parte crucial do trabalho futuro é a análise de fluxos completos, desde a recolha de dados até a sua preservação, numa situação em que os investigadores estão munidos com as ferramentas e infraestruturas adaptadas às suas necessidades. São necessárias soluções para problemas como a interoperabilidade tanto dentro das áreas quanto entre elas, visibilidade de resultados ao nível das instituições, visibilidade em repositórios internacionais, publicação de metadados, gestão das ligações entre os diferentes resultados de investigação. Para cada um deles vai ser necessário um esforço considerável, e é de esperar que as soluções gerais multidisciplinares tenham de ser avaliadas criticamente em cada domínio de investigação.

## REFERÊNCIAS

- AMORIM, R. C. et al. Efficient delivery of forecasts to a nautical sports mobile application with semantic data services. In: INTERNATIONAL CONFERENCE ON COMPUTER SCIENCE AND SOFTWARE ENGINEERING C3S2E, 9., 2016, Porto. **Proceeding...** Porto: Instituto Superior de Engenharia do Porto, 2016. p. 7-12.
- BRASE, J.; SENS, I.; LAUTENSCHLAGER, M. The tenth anniversary of assigning DOI names to scientific data and a five year history of DataCite. **D-Lib Magazine**, v. 21, n. 1-2, 2015.
- CAPLAN, P. **Understanding PREMIS**: an overview of the PREMIS data dictionary for preservation metadata. [Washington]: Library of Congress, 2009.
- CASTRO, J. A. et al. Ontologies for research data description: a design process applied to vehicle simulation. In: METADATA AND SEMANTICS RESEARCH CONFERENCE, 9., 2015, Manchester, UK. **Proceedings...** Manchester, UK: Springer, 2015. p. 348-354.
- CASTRO, J. A.; SILVA, J. R. da; RIBEIRO, C. Creating lightweight ontologies for dataset description: practical applications in a cross-domain research data management workflow. In: IEEE/ACM JOINT CONFERENCE ON DIGITAL LIBRARIES, 2014, London. **Proceedings...** 2014. Disponível em: <<http://ieeexplore.ieee.org/document/6970185/?reload=true>>. Acesso em: 22 jul. 2017.
- DATA ASSET FRAMEWORK. **Four steps to effective data management**. 2016. Disponível em: <<http://www.data-audit.eu/>>. Acesso em: 22 jul. 2017.
- DOTY, J. et al. Learning to curate. **Bulletin of the Association for Information Science and Technology**, v. 40, n. 6, 2015.
- EUROPEAN COMMISSION JOINT RESEARCH CENTRE. **INSPIRE Data Specification on Bio-geographical Regions**: technical guidelines. 2013. Disponível em: <<http://inspire.ec.europa.eu/id/document/tg/br>>. Acesso em: 22 jul. 2017.
- EUROPEAN STRATEGY FORUM ON RESEARCH INFRASTRUCTURES. **Strategy report on research infrastructures**. 2016. Disponível em: <[https://ec.europa.eu/research/infrastructures/pdf/esfri/esfri\\_roadmap/esfri\\_roadmap\\_2016\\_full.pdf](https://ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/esfri_roadmap_2016_full.pdf)>. Acesso em: 22 jul. 2017.
- FANIEL, I. M.; YAKEL, E. Significant properties as contextual metadata. **Journal of Library Metadata**, v. 11, n. 3/4, p.155-165, 2011.
- GOMES, F. et al. WindS@UP: the e-Science platform for WindScanner.eu. **Journal of Physics: Conference Series**, v. 524, p. 1-10, 2014. Disponível em: <<http://iopscience.iop.org/article/10.1088/1742-6596/524/1/012006/pdf>>. Acesso em: 22 jul. 2017.
- LECARPENTIER, D.; MICHELINI, A.; WITTENBURG, P. The building of the EUDAT cross-disciplinary data infrastructure. **Geophysical Research Abstracts**, v. 15, 2013. Disponível em: <<http://meetingorganizer.copernicus.org/EGU2013/EGU2013-7202-1.pdf>>. Acesso em: 22 jul. 2017.
- MANGHI, P. et al. An infrastructure for managing EC funded research output: the OpenAIRE Project. **The Grey Journal (TGJ): an International Journal on Grey Literature**, v. 6, n. 1, 2010.

MAYERNIK, M. S. **Metadata realities for cyberinfrastructure**: data authors as metadata creators. 2011. Dissertation (Doctor of Philosophy in Information Studies) – University of California, Los Angeles, 2011. Disponível em: <<http://dx.doi.org/10.2139/ssrn.2042653>>. Acesso em: 22 jul. 2017.

NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. **Nucleic Acids Research**, v. 41, 2013. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531099/>>. Acesso em: 22 jul. 2017.

RIBEIRO, C. et al. Motivators and deterrents for data description and publication: preliminary results. **Springer Lecture Notes in Computer Science**, v. 9416, 2015. p. 512-516.

RIBEIRO, C.; FERNANDES, M. E. M. Data curation at U.Porto: identifying current practices across disciplinary domains. **IASSIST Quarterly**, v. 35, n. 4, p.14-17, 2011. p.14-17.

SILVA, J. R. da; RIBEIRO, C.; LOPES, J. C. UPData: a data curation experiment at U. Porto using DSpace. In: INTERNATIONAL CONFERENCE ON PRESERVATION OF DIGITAL OBJECTS, 8., 2011, Singapore. **Proceedings...** Singapore, 2011.

SILVA, J. R. da; RIBEIRO, C.; LOPES, J. C. Usage-Driven Dublin Core descriptor selection: a case study using the Dendro Platform for research dataset description. **Springer Lecture Notes in Computer Science**, v. 9819, 2016.

# DSPACE CRIS

Lucas Rodrigues da Costa  
Milton Shintaku  
Lucas Ângelo da Silveira





## INTRODUÇÃO

O processo relacionado à geração do conhecimento científico, centrado nas pesquisas, possui diversas atividades envolvendo vários elementos. Tal processo é comumente conhecido como comunicação científica e tem sido influenciado pela tecnologia e a evolução social da comunidade de pesquisadores. Dessa forma, o processo de geração do conhecimento científico requer estudos constantes devido sua dinamicidade no cenário atual das tecnologias que atendem a ciência.

Em seu modelo de comunicação científica, Bjork (2007) apresenta alguns elementos que podem nortear o desenvolvimento de ferramentas voltadas à comunicação científica. O autor destaca que o modelo não apenas atualiza a representação do processo, mas também apresenta elementos que não são comumente observados, mesmo fazendo parte constante do processo.

No diagrama principal do modelo, apresentado na Figura 1, nota-se a visão social da ciência vista pelo autor, no qual um dos resultados de pesquisa é a melhoria da qualidade de vida. Essa é uma visão mais moderna da ciência, que se torna parte dos processos humanos na sua evolução social, que tem como base a melhoria das condições atuais e não apenas a criação de conhecimentos.

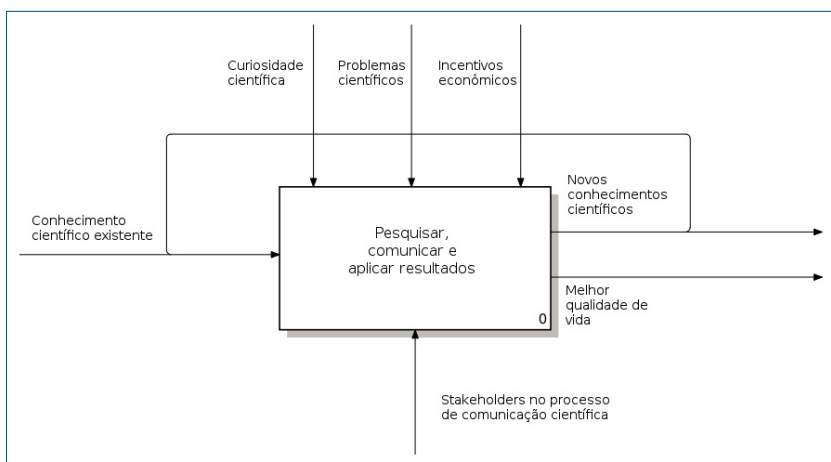


Figura 1 – Processo de pesquisar, comunicar e aplicar resultados

Fonte: Traduzido de Bjork (2007).



Outro ponto que se destaca são as motivações que geram uma pesquisa, na qual aparecem os tradicionais motivos como curiosidade científica e necessidade de solucionar problemas científicos. Vale ressaltar que o elemento que apresenta novidade está relacionado a questões econômicas, ou seja, a ciência sendo incentivada por questões econômicas. Esse ponto coloca a economia como parte do processo científico, tornando-o mais abrangente e complexo.

A existência de *stakeholders* no processo de comunicação científica não é uma ideia nova, mas o mapeamento desse elemento em cada etapa torna-se crucial ao entendimento dos novos elementos que atuam na pesquisa. Por exemplo, Lievrouw e Carley (1990) incluem a popularização da ciência como um elemento importante na comunicação científica, na qual a opinião pública pode atuar como um *stakeholder* nas pesquisas. Ações sociais, pesquisas genéticas humanas e células tronco são exemplos dessa afirmação.

Nesse contexto, nota-se que o cenário da comunicação científica compõe-se de vários elementos, forças, motivações e outros componentes que atuam de forma diferenciada influenciando sua execução. Isso requer tecnologias que atendam a toda essa complexidade, mesmo que atuem em uma determinada tarefa, contemplando todos os elementos envolvidos.

Assim, será apresentada a ferramenta livre DSpace versão para o *Current Research Information System* (CRIS), que estende as funcionalidades da versão padrão. Essa ferramenta procura integrar novas funcionalidades, atendendo a outros elementos incluídos no papel ofertado pelos repositórios científicos.

## **CURRENT RESEARCH INFORMATION SYSTEM (CRIS)**

Com a informatização e a evolução das ciências, a quantidade de informações científicas disponíveis apresenta um desafio na avaliação do que é ou não relevante. Em muitos casos, documentos e informações relacionadas estão descontextualizados ou espalhados por vários sistemas de informação. Nem mesmo ferramentas de busca potentes como o Google resolvem todos os problemas voltados à recuperação de informação.

Para apoiar a recuperação de informação de pesquisa, como afirma Lopatenko (2001), muitas instituições estão apostando na criação de sistemas CRIS, com adição de ferramentas de web semântica como as que lidam com ontologias. Nesse sentido, o referido autor advoga pela criação de sistemas CRIS que utilizam web semântica, de forma que pesquisadores e instituições de ensino e pesquisa registrem o conhecimento no intuito de melhorar a recuperação da informação.

Esse ponto apresenta uma das finalidades dos sistemas CRIS, ou seja, a recuperação contextualizada da informação. Entretanto, um sistema de informação



sobre pesquisas correntes se apresenta maior e mais complexo, envolvendo informações consolidadas sobre os diversos elementos que envolvem a pesquisa, contemplando ferramentas de buscas que utilizem funcionalidades baseadas em web semântica.

Johansson e Ottosson (2012) diferem os sistemas CRIS dos Repositórios Institucionais (RI) tradicionais. Os autores destacam que os sistemas CRIS contêm informações sobre as pesquisas em andamento, enquanto os RIs cuidam apenas das publicações em seus textos integrais. Isso reforça o uso de sistemas CRIS na recuperação da informação, visto que são mais abrangentes quando comparados com RIs de forma geral.

Os sistemas CRIS mostram-se mais abrangentes, principalmente por conter as informações sobre as pesquisas correntes e não apenas aos resultados em forma de publicações, como ocorre nos sistemas de informações sobre pesquisas atuais. Com isso, revela-se uma opção para instituições que queiram disseminar informações mais contextualizadas sobre suas pesquisas correntes e passadas.

Em contrapartida, como apresentam Moreira, Cunha e Macedo (2015), há sistemas CRIS e ecossistemas CRIS. Nesse sentido, os autores apresentam o PTCRIS, um ecossistema nacional, federado por sistemas locais. Assim, deve-se distinguir os ecossistemas CRIS compostos por vários sistemas independentes em forma de federação, dos sistemas CRIS que agregam informações de pesquisas correntes de uma instituição.

Os sistemas CRIS, além de disseminar informações sobre pesquisas correntes, podem ser utilizados para diversos fins, tanto que, Bjork (2007) inclui o processo de depósito de metadados nesses sistemas na etapa de Avaliação Prévia dos Candidatos (Figura 2). Segundo esse autor, as informações inseridas nos sistemas CRIS de uma universidade podem alimentar processos de avaliação do desenvolvimento de pesquisa, a fim de verificar o seu desempenho. Com isso, órgãos de fomento e universidades podem avaliar os candidatos a novos pedidos de financiamentos com informações das pesquisas correntes. Dessa forma, os sistemas CRIS são vistos como apoiadores a tomada de decisão, principalmente para órgãos de fomento de pesquisas, fornecendo dados de pesquisas correntes e passadas de instituições e pesquisadores.

Nessa mesma linha, Sheppard (2010) apresenta o modelo estratégico do sistema CRIS (Figura 3), destacando outros relacionamentos necessários para a gestão de projetos correntes. Os resultados de pesquisa são importantes e podem ser verificados pelas publicações; entretanto, informações como as dos projetos e financiamentos tornam-se essenciais aos sistemas CRIS. Assim, pode-se ofertar informações consolidadas a diversos tipos de usuários, com maior ênfase na gestão da informação, podendo ser utilizada para a tomada de decisões em vários segmentos da sociedade.

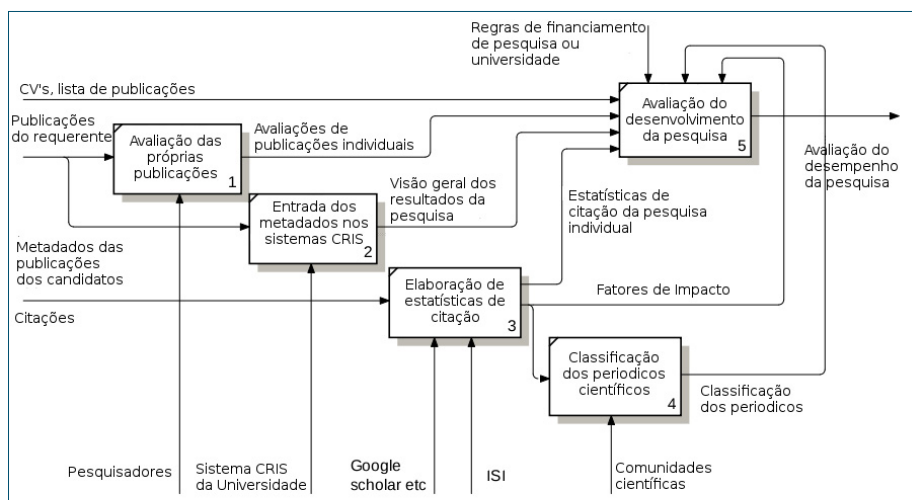


Figura 2 – Diagrama do processo de avaliação prévia dos candidatos

Fonte: Traduzido de Bjork (2007).

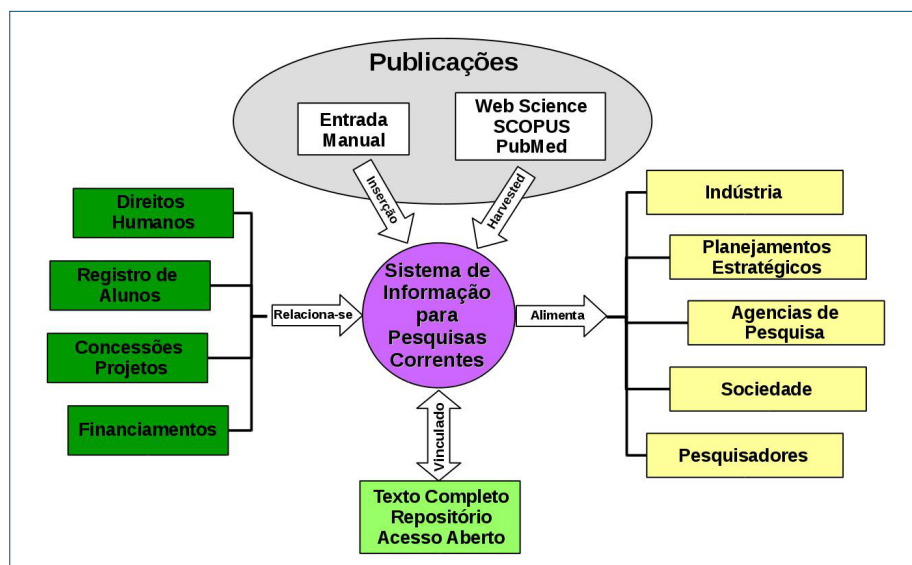


Figura 3 – Modelo estratégico do Sistema CRIS

Fonte: Traduzido de Sheppard (2010).

Para tanto, os sistemas CRIS requerem modelo de metadados mais complexos, sendo adotado o *Common European Research Information Format* (CERIF). O CERIF possibilita a descrição de várias entidades, como mostra a Figura 4. Com esse esquema de metadados, pode-se descrever diversas entidades de forma relacionada,



gistrados como facilitadores do acesso. Assim, as principais atividades dos repositórios consistem nos processos de entrada, organização desses documentos e disponibilização controlada do seu acesso. Dessa forma, o DSpace-CRIS mantém essas funcionalidades, agregando outras para se adaptar aos preceitos dos sistemas CRIS.

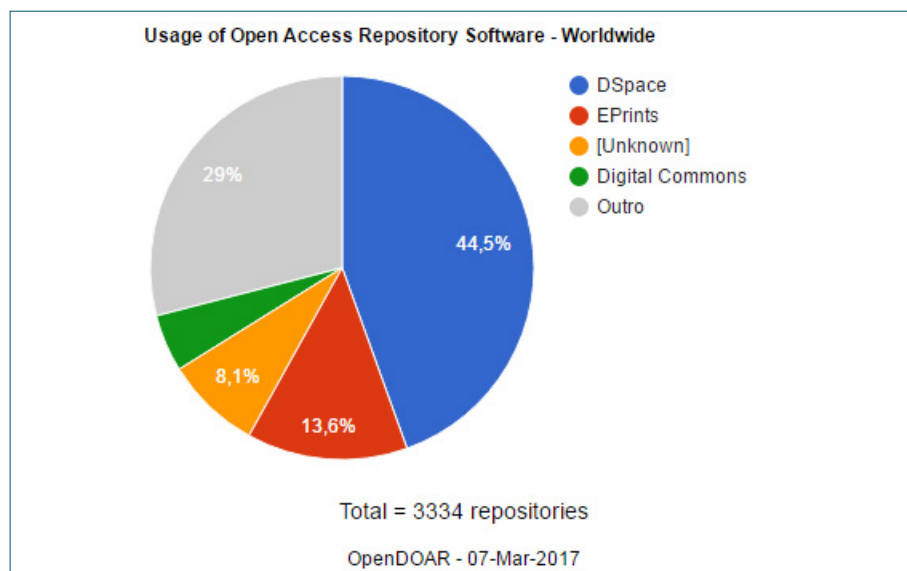


Figura 5 – Distribuição dos repositórios de acesso livre  
Fonte: OpenDOAR (2017).

Nesse sentido, Silveira, Shintaku e Bollini (2016) relatam que o DSpace-CRIS expande as funcionalidades do DSpace padrão, agregando funcionalidades relacionadas a pessoas, projetos e organizações, como mostra a Figura 6. Palmer et al. (2014) revelam que a adoção do DSpace-CRIS pela Universidade de Hong Kong, utilizando o esquema de metadados CERIF, melhorou a obtenção de informações sobre os relacionamentos entre pesquisadores e avaliações das pesquisas.

Assim, pode-se estabelecer que o DSpace-CRIS está firmado em quatro pilares que se inter-relacionam. A Figura 7 apresenta a estrutura do DSpace-CRIS com seus respectivos pilares. Um pesquisador (**pessoa**) está vinculado a um instituto de pesquisa ou universidade (**organização**), possuindo um projeto de pesquisa (**projeto**) que produziu um artigo (**publicado**). O projeto de pesquisa (**projeto**), por sua vez, é financiado por um órgão de fomento (**organização**), produzindo documentação científica (**publicação**).

Obviamente, os relacionamentos internos são possíveis e comuns, pesquisadores (**pessoa**) possui relação com outros pesquisadores (**pessoa**) e outros envolvidos na pesquisa (**pessoa**), assim como, **publicações** se relacionam tematicamente, universidades (**organização**) se relaciona com órgão de fomentos e institutos de pesquisa (**organização**) e assim por diante. Com isso, o grande diferencial do DSpace CRIS é possibilitar o tratamento de informações sobre esses pilares e relacioná-los.



Figura 6 – Estrutura do DSpace CRIS  
Fonte: Silveira, Shintaku e Bollini (2016).

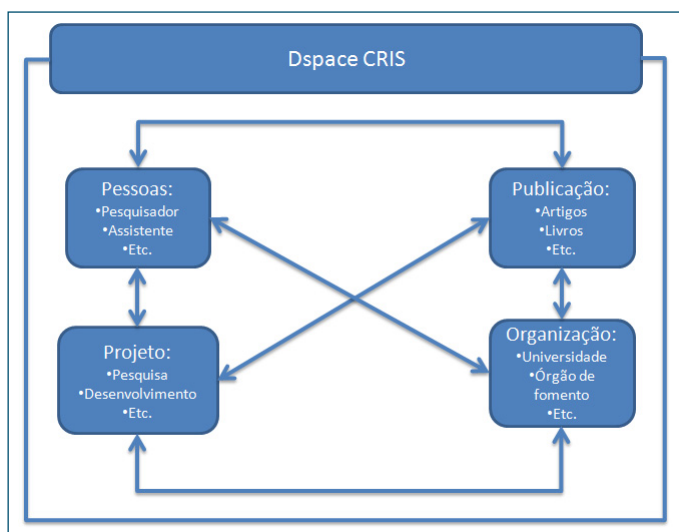


Figura 7 – Estrutura de relacionamento do DSpace-CRIS  
Fonte: Autoria própria (2017).

Esses pilares recebem destaque no DSpace CRIS, à medida que estão presentes na barra principal de navegação, Figura 8, que apresenta as seguintes opções: Comunidades e Coleções, Saída de pesquisas (**publicações**), pesquisadores (**pessoas**), **organizações** e **projetos**. Cada uma dessas opções possibilita recuperar as informações relacionadas, inseridas em outro processo que requer permissões, entre outros pontos.

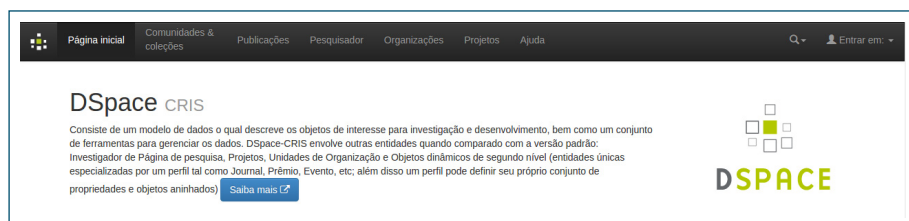


Figura 8 – Detalhe da página inicial do DSpace CRIS

Fonte: Autoria própria (2017).

Cada uma das opções do menu principal de navegação remete a uma parte do sistema, de forma a recuperar as informações. Com isso, o DSpace CRIS oferta uma seção dentro do sistema para cada um dos pilares, possibilitando apresentar as suas informações e as pertinentes e suas relações. Assim, possibilita a navegação pelas informações geridas pelo DSpace CRIS.

Os exemplos e informações apresentados neste capítulo são resultados de pesquisas efetuadas no âmbito do projeto acordado entre o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) e a Secretaria Nacional de Juventude (SNJ). Assim, nomes dos elementos de informação como as comunidades e coleções, publicações, pesquisadores e outros terão relação com o projeto, oferecendo a visão aplicada do DSpace-CRIS, com todas as relações.

## COMUNIDADES E COLEÇÕES

A forma de organização do DSpace CRIS segue a mesma do DSpace padrão, com a hierarquia formada por comunidades, subcomunidades, coleções, itens e arquivos, sendo que as três primeiras são utilizadas para criar uma forma de organização do repositório. As comunidades podem representar grandes temas tratados pelo instituto, universidade ou organização; subcomunidades são refinamentos da comunidade, enquanto as coleções podem representar conjuntos de objetos digitais.

A página das comunidades e coleções lista a estrutura criada no DSpace CRIS, como mostra em detalhes a Figura 9. A lista de comunidade, subcomunida-

des e coleções é apresentada em ordem alfabética, de modo que, caso se queira dar uma ordem diferente, recomenda-se utilizar números como parte do nome da comunidade, subcomunidade e coleção. As informações apresentadas, inclusive o seu ícone, foram inseridas no momento de criação da comunidade, subcomunidade e coleção. Pode-se, no entanto, alterar essas informações se for preciso.



Figura 9 – Detalhe da página da opção comunidade e coleção

Fonte: Autoria própria (2017).

No exemplo demonstrado, a comunidade apresenta um tema maior, **Tecnologias para Gestão da Informação**, que é refinado na subcomunidade **Ferramentas para Gestão de Bibliotecas**, Figura 9. Assim, a comunidade é um tema e as subcomunidades subtemas. As coleções, neste caso, são conjuntos de documentos relacionados por alguma característica. No exemplo, a coleção é o conjunto de documentos sobre o Koha, software livre para gestão de bibliotecas.

Tanto o ícone da comunidade, subcomunidade e coleção, quanto o nome na lista apresentada, torna-se um link para a página de descrição da comunidade, subcomunidade e coleção. A Figura 10 mostra a página da comunidade **Tecnologia para Gestão da Informação**, com as suas informações, documentos mais recentes (Letra C), submenu de navegação pelos documentos da comunidade (Letra B) e suas subcomunidades (Letra D).

Assim, as comunidades, subcomunidades e coleções servem aos usuários que preferem navegar pelo repositório à procura de documentos. Pode-se navegar pela estrutura do Repositório em busca de informações, depositadas conforme a organização disposta pelas comunidades, subcomunidades e coleções. A comunidade apresentada como exemplo na Figura 10 revela que esse elemento de informação está focado nos documentos, na forma de organização do acervo depositado no repositório, pois todas as suas funcionalidades voltam-se para a recuperação desses documentos.



Figura 10 – Informação sobre a comunidade

Fonte: Autoria própria (2017).

## PUBLICAÇÕES

As publicações já existiam no DSpace padrão, mas como parte da hierarquia informacional do DSpace, sendo denominada de Item (conjunto formado pelos metadados e arquivos). Com o DSpace-CRIS, as publicações ganham destaque, formando um dos pilares desse sistema de informação, ganhando páginas que apresentam as publicações de diversas formas.

Ao acessar a página de publicações disponível no menu principal do DSpace-CRIS, o sistema remete à página que possibilita acessar documentos pertencentes ao acervo. Com isso, pode-se explorar as publicações de várias formas, facilitando a recuperação. Expandem-se as possibilidades de recuperação das publicações mantidas pelo DSpace, possibilitando a navegação por título, autores e outras formas já existentes, adicionando os depósitos mais recentes e as publicações mais acessadas. Facetamentos são ofertados para ajudar a refinar a navegação.

Entretanto, pode-se dizer que a grande alteração ofertada pelo DSpace-CRIS está na apresentação das publicações. No exemplo da Figura 11, extraído do DSpace-CRIS pertencente ao projeto EuroCRIS, as informações da publicação são apresentadas como no DSpace padrão, apresentado metadados como autores, título, resumos e outros. A novidade fica por conta das relações com alguns indicadores.



A página da publicação apresenta links para indexadores como o *Web of Science* e *Scopus*, possibilitando verificar informações sobre as citações. Pode-se, também, verificar os indicadores de visualização e download, com representação gráfica baseada em georreferenciamento, da mesma forma que liga ao Google Acadêmico e traz informações sobre altimetria, possibilitando acessar as informações altimétricas da publicação no site do *Altmetrics*.





<b>Title:</b>	The Quest for Research Information	<div>  <b>Page view(s)</b>  <b>923</b> ⓘ  <small>checked on Mar 26, 2017</small> </div> <div>  <b>Download(s)</b>  <b>580</b> ⓘ  <small>checked on Mar 26, 2017</small> </div> <div>  <b>Google Scholar™</b>  <b>Check</b> </div> <div>  <b>Altmetric</b> </div>
<b>Authors:</b>	Blümel, Ina ⓘ Dietze, Stefan ⓘ Heller, Lambert ⓘ Jäschke, Robert ⓘ Mehberg, Martin ⓘ	
<b>Keywords:</b>	research information management linked open data web crawling information extraction	
<b>Issue Date:</b>	13-May-2014	
<b>Publisher:</b>	euroCRIS	
<b>Source:</b>	"Managing Data-Intensive Science: the Role of Research Information Systems in Realising the Digital Agenda". Proceedings of the 12th International Conference on Current Research Information Systems (2014) Procedia Computer Science 33: 253-260 (2014)	
<b>Series/Report no.:</b>	CRIS2014: 12th International Conference on Current Research Information Systems (Rome, May 13-15, 2014)	
<b>Conference:</b>	CRIS2014 Conference ⓘ	
<b>Abstract:</b>	Research information, i.e., data about research projects, organisations, researchers or research outputs such as publications or patents, is spread across the web, usually residing on institutional and personal web pages or in semi-open databases and information systems. While there exists a wealth of unstructured information, the limited amounts of structured data often are exposed following proprietary or less-established schemas and interfaces. Therefore, a holistic view on research information across organisational and national boundaries is not feasible and information is inconsistent and incomplete. On the other hand, web crawling and information extraction techniques have matured throughout the last decade, allowing for automated approaches of harvesting, extracting and consolidating research information into a more coherent knowledge graph. In particular the Linked Data community has provided a range of techniques, schemas and vocabularies which allow to represent and interlink research information in a more coherent manner. In this work, we give an overview of the current state of the art in research information sharing on the web and present initial ideas towards a more holistic approach for bootstrapping research information from available web sources.	

Figura 11 – Página do registro no EuroCRIS<sup>3</sup>  
 Fonte: EuroCRIS (2017).

As novas funcionalidades disponíveis no DSpace-CRIS revelam a forte integração desse sistema de informação com outros sistemas, alinhado às tendências de compartilhamento de informações. Assim, possibilita obter indicadores relacionados à publicação, oriundos de diversas fontes, tanto as tradicionais como a *Web of Science*, quanto as mais novas como a *Altmetrics*, abrangendo informações relevantes quanto ao acesso e uso dos documentos do acervo.

Deve-se, no entanto, destacar que o link aparece somente se houver dados. Caso um documento não esteja presente na base do *Scopus* ou *Web of Science*, não aparecerá o link para esses indexadores. Provavelmente, o link mais comum em todas as apresentações de publicações talvez seja o Google Acadêmico, visto que publicações hospedadas em repositórios geralmente são indexadas por esse sistema.

<sup>3</sup> Disponível em: <<http://dspacecris.eurocris.org/handle/11366/194>>. Acesso em: 22 jul. 2017.

## PESSOAS

Pessoas ou pesquisadores no DSpace-CRIS difere dos usuários do repositório, pois os pesquisadores são unidades de informação, enquanto os usuários executam atividades no repositório. No DSpace padrão não existe a figura do pesquisador, tendo apenas os usuários, que são utilizados para controle de acesso ao conteúdo e serviços.

Pesquisadores são unidades de informação com página própria (Figura 12), e formam no DSpace-CRIS uma base de controle de autoridade, o que possibilita que um mesmo pesquisador seja identificado por variantes, melhorando as formas de recuperação da informação pelo nome do pesquisador, da mesma forma que possibilita a normalização das entradas de autoria, visto que basta selecionar o nome constante na base de autoria.

**Costa L. R.** [Network Lab](#) [Visualizar estatísticas](#) [Email de alerta](#) [RSS Feed](#)


Profile Indicators Other

Publications (Outros) [Mostrar/Ocultar Filtros](#)

Resultados 1-2 of 2 (Tempo de busca: 0.004 segundos).

Data do envio	Título	Autor(s)
1 2016	REVISTA DE JUVENTUDE E POLITICAS PUBLICAS: UM RELATO DE EXPERIÊNCIA DE CRIAÇÃO E DESENVOLVIMENTO	Brito, R. F.  Shintaku, M.  Jesus, Jaqueline R. de  Macedo, D. J.  SCHIESSEL, I. T.  CASTRO, P. P.  SANTOS, P. R.  DURADO, A. M. C. S. F.  Silveira, L. A.  Costa L. R.
2 2016	PROBLEMAS NA UTILIZAÇÃO DO OPEN JOURNAL SYSTEM (OJS/IBER)	Brito, Ronnie Fagundes  Shintaku, M.  Jesus, Jaqueline R. de  Macedo, D. J.  SCHIESSEL, I. T.  CASTRO, P. P.  SANTOS, P. R.  DURADO, A. M. C. S. F.  Silveira, L. A.  Costa L. R.

Profile



**Credit Name** Costa L. R. **Full Name** Lucas Rodrigues Costa

**Versuscular Name** Costa L. R.

**Main Affiliation** [IBICT](#)

**Working groups** [Coordenação de Atuação, Gestão e Aplicação de Tecnologia](#)

**Personal site** <http://lattes.cnpq.br/3133273170328412>

**Email** [lucasrodrigues@ibict.br](mailto:lucasrodrigues@ibict.br)

**ORCID** [0000-0002-0973-4866](https://orcid.org/0000-0002-0973-4866)

**Biography**  
Lucas R. Costa received his B.Sc. and M.Sc. degrees in Computer Science (2012, 2016) from the University of Brasília, Brazil. He is currently working toward a Ph.D. degree in Computer Science at University of Brasília. His research interest includes Computer Networks, Optical Networks and Traffic Engineering with emphasis on network control and management.

Figura 12 – Página de registro dos pesquisadores no DSpace-CRIS COTEC<sup>4</sup>  
Fonte: DSpaceCRIS COTEC (2017).

4 Disponível em: <<https://coteccris.ibict.br/jspui/cris/rp/00004>>. Acesso em: 22 jul. 2017.

Na página do pesquisador são apresentadas as informações gerais e as publicações mais recentes, com a opção por navegar pelas publicações deste pesquisador. Possivelmente a grande novidade do DSpace-CRIS em relação aos pesquisadores é a integração com a base de dados *Open Researcher and Contributor Identification* (ORCID). Outro ponto de destaque é a possibilidade de geração gráfica das relações entre os pesquisadores, baseadas nas co-autorias, como mostra a Figura 13. Para acessar o gráfico, basta clicar na aba **Network Lab**.

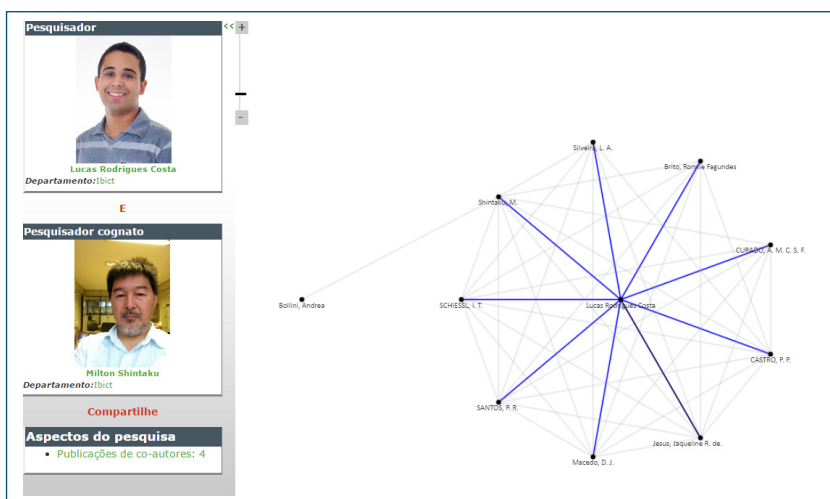


Figura 13 – Página do resultado de relação entre os pesquisadores no DSpace-CRIS COTEC<sup>5</sup>

Fonte: DSpaceCRIS COTEC (2017).

Na página do pesquisador, a tônica está na integração com outras bases de dados, como o ORCID e *Scopus*. Cabe destacar que o pesquisador pode sincronizar o DSpace-CRIS com a sua página ORCID, enviando as informações sobre as suas publicações, ou seja, o DSpace-CRIS pode se tornar uma fonte de informação para o sistema ORCID.

## ORGANIZAÇÕES

A organização é outra unidade de informação existente no DSpace-CRIS, que não era contemplado no DSpace padrão, possibilitando a descrição e agregação de informação sobre um instituto de pesquisa, universidade, órgão de fomento e outras entidades relacionadas à organização. Com isso, oferta aos usuários a recuperação de informação consolidada.

<sup>5</sup> Disponível em: <<https://coteccris.ibict.br/jspui/cris/network/rp00004>>. Acesso em: 22 jul. 2017.

Com a organização, pode-se relacionar projetos, publicações e pesquisadores com uma determinada entidade organizacional, possuindo uma página, como mostra a Figura 14. Da mesma forma, pode-se criar as suborganizações, possibilitando criar uma estrutura hierárquica que refina as informações.

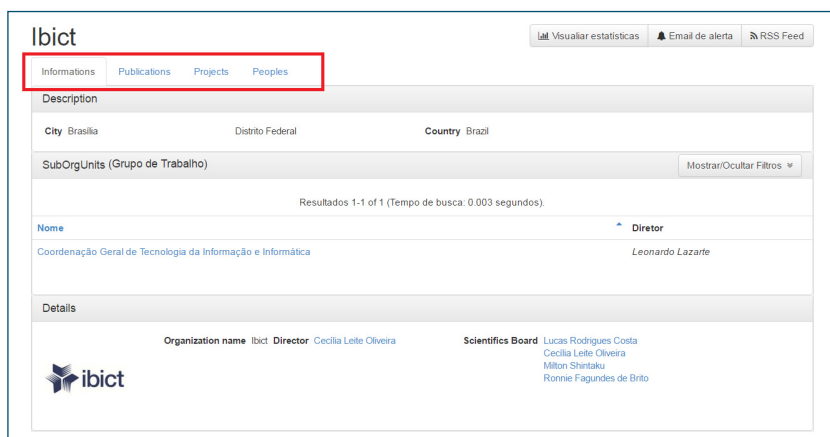


Figura 14 – Página de registro das organizações do DSpace-CRIS COTEC<sup>6</sup>

Fonte: DSpace-CRIS COTEC (2017).

Para cada página de organização ou suborganização, têm-se abas que permitem verificar informações classificadas, no caso, publicações, projetos e pesquisadores. Pode-se criar novas abas caso seja necessário, mas isso requer conhecimentos sobre o sistema DSpace-CRIS, de forma a obter ou reunir as informações pertinentes à nova aba.

Organizações tornam-se importantes, principalmente, para verificar as relações interinstitucionais das pesquisas, à medida que pode-se ter pesquisas ou publicações (como resultados de pesquisa) que envolvam mais de uma instituição. Assim, pode-se cadastrar todas as entidades que se relacionam a uma pesquisa para construir a rede necessária ao entendimento global das pesquisas.

## PROJETOS

Projetos de pesquisas correntes, em muitos casos, são difíceis de gerenciar, pois em fase inicial não possuem publicações. Entretanto, depois de certo tempo ou de o projeto ter acabado, as suas informações de publicação e pesquisadores podem ser valiosas para os gestores. Assim, os projetos podem ser um elemento de informação no DSpace-CRIS, com página própria, como mostrado na Figura 15.

<sup>6</sup> Disponível em: <<https://coteccris.ibict.br/jspui/cris/ou/ou00003>>. Acesso em: 22 jul. 2017.



Figura 15 – Página de registro de projeto do DSpace-CRIS COTEC<sup>7</sup>

Fonte: DSpaceCRIS COTEC (2017).

Neste exemplo, o projeto é de integração de sistemas, ou seja, cada projeto possui sua própria página, possibilitando recuperar informações consolidadas, como neste caso as informações gerais, descrição, publicações e fomento. As abas, em destaque na Figura 15, podem ser criadas, caso haja conhecimento do sistema e da oferta de informações.

Provavelmente, uma das grandes vantagens desse elemento de informação é a possibilidade de agregação de todas as publicações de um determinado projeto em uma mesma página, visto que traz os registros dos resultados obtidos neste projeto. Nos repositórios padrão DSpace tradicionais, essa informação não existe, a não ser que tenha um metadados exclusivo para isso, dificultando, muitas vezes, o levantamento dos resultados de pesquisa.

Neste contexto, é possível reunir informações de um projeto em repositórios desde a sua criação, com as informações iniciais de descrição, coordenador, contato, duração e outros. Outra grande vantagem é a possibilidade de vinculação dos pesquisadores aos projetos, assim como, as organizações envolvidas. A tônica dos sistemas CRIS são as pesquisas correntes, muitas vezes ligadas aos seus projetos.

## CONSIDERAÇÕES FINAIS

Os sistemas de informações tornam-se mais abrangentes, à medida que a atividade à qual o sistema apoia requer mais serviços informacionais. Assim, os sistemas informatizados estão em constante aperfeiçoamento, revendo os módulos existentes e adicionando novos, com a oferta de novas funcionalidades.

<sup>7</sup> Disponível em: <<https://coteccris.ibict.br/jspui/cris/project/pj00001>>. Acesso em: 22 jul. 2017.

Para os softwares livres, o cenário se torna mais diversificado, visto que uma ferramenta pode ter distribuições diferenciadas, onde cada distribuição possui características próprias que atendem a determinadas necessidades. Esse é o caso, por exemplo, do sistema operacional Linux, que atualmente se compõe de uma vasta família de distribuição.

Esse também parece ser o caso do DSpace, que ofertou a distribuição do DSpace-CRIS pelo Cinecca. Assim, dá-se a oportunidade aos usuários de selecionar a distribuição que melhor atende à instituição. O DSpace-CRIS torna-se uma opção aos repositórios que desejam incrementar as suas informações, adicionando outros elementos de informação, abrangendo informações sobre pesquisadores, projetos e organizações.

Com isso, atende-se às necessidades emergentes de se ter informações consolidadas das pesquisas correntes. O DSpace-CRIS oferta-se aos gestores de projetos, agências de fomento, avaliadores e outros usuários. Sua principal característica é o relacionamento desses elementos de informação para a formação de novas informações, podendo inclusive apoiar decisões.

A adoção do DSpace CRIS pelos projetos EuroCRIS, como repositório do projeto, reforça a opção por essa distribuição do DSpace, mesmo que ainda não seja muito disseminado entre os usuários de repositórios. O pioneirismo da Universidade de Hong Kong, com o primeiro DSpace-CRIS, parece estar dando frutos, ainda mais que essa ferramenta está sendo divulgada pela comunidade e mantida pela 4Science.

Por fim, o DSpace-CRIS contribui para a evolução do DSpace e seus usuários, com o intuito de apoiar a sua liderança de uso. Da mesma forma, com este estudo espera-se contribuir para a discussão sobre repositórios digitais, apresentando esta ferramenta abrangente que transcende a disseminação de resultados de pesquisa e mantém a familiaridade existente do já conhecido DSpace.

## REFERÊNCIAS

BJÖRK, B. A model of scientific communication as a global distributed information system. *Information Research*, v. 12, n. 2, 2007. Disponível em: <<http://InformationR.net/ir/12-2/paper307.html>>. Acesso em: 09 abr. 2017.

DSPACECRIS COTEC. **Repositório da Coordenação de Articulação, Geração e Aplicação de Tecnologia**. Disponível em: <<http://coteccris.ibict.br/>> Acesso em: 07 abr. 2017.

EUROCRIS. **EuroCRIS DSpace CRIS digital reposit**. Disponível em: <<http://dspacecris.eurocris.org/>>. Acesso em: 09 abr. 2017.

JOHANSSON, Å.; OTTOSSON, M. O. A national current research information system for Sweden. In: INTERNATIONAL CONFERENCE ON CURRENT RESEARCH INFORMATION SYSTEMS, 11., 2012, Prague, Czech Republic. **Proceedings...** Prague, Czech Republic, 2012. p. 67-71. Disponível em: <[http://dspacecris.eurocris.org/bitstream/11366/103/1/CRIS2012\\_Johansson\\_SwedishCRIS.pdf](http://dspacecris.eurocris.org/bitstream/11366/103/1/CRIS2012_Johansson_SwedishCRIS.pdf)>. Acesso em: 07 abr. 2017.

LIEVROUW, L. A.; CARLEY, K. Changing patterns of communication among scientists in an era of "telescience". **Technology in Society**, v. 12, n. 4, p. 457-477, 1990. Disponível em: <[http://casos.cs.cmu.edu/publications/papers/lievrouw\\_1990\\_changingpatterns.PDF](http://casos.cs.cmu.edu/publications/papers/lievrouw_1990_changingpatterns.PDF)>. Acesso em: 09 abr. 2017.

LOPATENKO, A. Information retrieval in current research information systems. **arXiv**, preprint cs/0110026, 2001. Disponível em: <<https://arxiv.org/ftp/cs/papers/0110/0110026.pdf>>. Acesso em: 07 abr. 2017.

MOREIRA, J. M.; CUNHA, A.; MACEDO, N. An ORCID based synchronization framework for a national CRIS ecosystem. **F1000Research**, v. 4, p. 1-17, 2015. Disponível em: <<https://f1000research.com/articles/4-181/v1#>>. Acesso em: 07 abr. 2017.

OpenDOAR. **Usage of open access repository software**: worldwide. Reino Unido: Universidade de Nottingham, 2017. Disponível em: <<http://www.opendoar.org/onechart.php?clID=&ctID=&rtID=&clID=&lID=&potID=&rSoftWareName=&search=&groupby=r.rSoftWareName&orderby=Tally%20DESC&charttype=pie&width=600&height=300&caption=Usage%20of%20Open%20Access%20Repository%20Software%20-%20Worldwide>>. Acesso em: 03 mar. 2017.

PALMER, D. T. et al. DSpace-CRIS@ HKU: achieving visibility with a CERIF compliant open source system. **Procedia Computer Science**, v. 33, p. 118-123, 2014. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1877050914008096>>. Acesso em: 09 abr. 2017.

SHEPPARD, N. Learning how to play nicely: repositories and CRIS. **Ariadne**, Loughborough, n. 64, jul. 2010. Disponível em: <<http://www.ariadne.ac.uk/issue64/wrn-repos-2010-05-rpt>>. Acesso em: 09 abr. 2017.

SILVEIRA, L. A.; SHINTAKU, M.; BOLINNI, A. **Guia de instalação do DSpace-CRIS**. Brasília: IBICT, 2016. Disponível em: <[http://livroaberto.ibict.br/bitstream/123456789/1066/2/IBICT\\_Guia\\_de\\_Instala%c3%a7%c3%a3o\\_DSpaceCRIS.pdf](http://livroaberto.ibict.br/bitstream/123456789/1066/2/IBICT_Guia_de_Instala%c3%a7%c3%a3o_DSpaceCRIS.pdf)>. Acesso em: 07 abr. 2017.







# ORGANIZADORES

(em ordem alfabética)

## **Clediane Guedes**

Graduada em Biblioteconomia pela UFRN. Mestranda em Gestão da Informação e do Conhecimento pela UFRN. Atua na área de Repositórios Digitais, com ênfase em repositório institucional, desenvolvido em Dspace.

## **Débora Koshiyama**

Graduada em Biblioteconomia pela Universidade Federal do Rio Grande do Norte (UFRN). Mestra em Design pela UFRN. Atua na área de Desenho Industrial, com ênfase em Desenho Industrial, atuando principalmente nos seguintes temas: design, design centrado no usuário, repositório institucional, divulgação científica e usabilidade.

## **Elisângela Moura**

Graduada em Biblioteconomia pela Universidade Federal do Rio Grande do Norte (UFRN). Atua na área de pesquisa de Ciência da Informação.

## **Emanuelle Torino**

Coordenadora do Comitê Gestor do Portal de Informação em Acesso Aberto da UTFPR (PIAA), do Repositório Institucional da UTFPR (RIUT) e do Repositório de Outras Coleções Abertas (ROCA), membro do Comitê Gestor do Portal de Periódicos Científicos da UTFPR (PERI). Coordenadora-Adjunta da Editora da Universidade Tecnológica Federal do Paraná (UTFPR). Graduada em Biblioteconomia pela Universidade Estadual de Londrina (UEL). Mestre em Gestão da Informação pela UEL. Atua nas linhas de pesquisa: gestão da informação e do conhecimento, conhecimento científico, acesso aberto, repositórios institucionais e periódicos científicos.

## **Fernando Vechiato**

Professor adjunto da Universidade Federal do Rio Grande do Norte (UFRN). Membro do corpo docente do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP). Graduado em Biblioteconomia pela UNESP. Mestre em Ciência da Informação pela UNESP. Doutor em Ciência da Informação pela UNESP. Atua na área de Ciência da Informação, atuando principalmente nos seguintes temas: encontrabilidade da informação, arquitetura da informação, usabilidade, acessibilidade, comportamento informacional, repositórios digitais.

## **Maria Aniolly Maia**

Graduada em Biblioteconomia pela Universidade Federal do Rio Grande do Norte (UFRN). Mestra em Design pela UFRN. Doutoranda em Gestão e Organização do Conhecimento pela Universidade de Minas Gerais (UFMG). Atua nas áreas de pesquisa: design de interfaces, usabilidade, arquitetura da informação e periódicos científicos.

## **Tércia Marques**

Graduada em Biblioteconomia pela Universidade Federal do Rio Grande do Norte (UFRN). Mestre em Educação pela UFRN. Atua nas linhas de pesquisa: estágio supervisionado, história da educação, conservação e preservação de acervos raros.

# AUTORES

(em ordem alfabética)

## Adriana Carla Silva de Oliveira

Professora da Universidade do Rio Grande do Norte (UFRN). Membro do corpo docente do Programa de Pós-Graduação em Direito da UFRN. Graduada em Biblioteconomia pela Universidade Federal da Paraíba (UFPB) e em Direito pela Universidade Potiguar (UnP). Mestre em Engenharia de Produção pela UFRN. Doutora em *Information Science* pela University of Tennessee (UTK) e em Ciência da Informação pela UFPB. Pós-Doutoranda em Direito pela UFRN. Atua nas linhas de pesquisa: ciência da informação, direito, tecnologia da informação e biblioteconomia. Membro do Grupo de Pesquisa Direito e Desenvolvimento do Departamento de Direito Processual e Propedêutica (DEPRO) da UFRN.

## Ana Alice Baptista

Professora da Universidade do Minho (UMinho). Graduada em Engenharia de Sistemas de Informação, Mestre em Informática e Doutora em Tecnologias e Sistemas de Informação todos pela UMinho. Atua nas linhas de pesquisa: metadados, *Linked Open Data* e movimento *open* em geral.

## Cristina Ribeiro

Professora Associada da Faculdade de Engenharia da Universidade do Porto (FEUP). Doutora em Informática pela Universidade Nova de Lisboa (UNL). Atua nas linhas de pesquisa: recuperação de informação, preservação digital e gestão de dados de investigação. Membro do Grupo de Pesquisa InfoLab da FEUP.

## Emanuelle Torino

Coordenadora do Comitê Gestor do Portal de Informação em Acesso Aberto da UTFPR (PIAA), do Repositório Institucional da UTFPR (RIUT) e do Repositório de Outras Coleções Abertas (ROCA), membro do Comitê Gestor do Portal de Periódicos Científicos da UTFPR (PERI). Coordenadora-Adjunta da Editora da Universidade Tecnológica Federal do Paraná (UTFPR). Graduada em Biblioteconomia pela Universidade Estadual de Londrina (UEL). Mestre em Gestão da Informação pela UEL. Atua nas linhas de pesquisa: gestão da informação e do conhecimento, conhecimento científico, acesso aberto, repositórios institucionais e periódicos científicos.

## Felipe Augusto Arakaki

Bibliotecário do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP). Graduado em Biblioteconomia pela Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP). Mestre e Doutorando em Ciência da Informação pela UNESP. Atua nas linhas de pesquisa: ciência da informação: biblioteca escolar, representação e organização da informação, catalogação, catalogação automatizada, metadados, interoperabilidade, padrões de metadados, *Dublin Core*, BIBFRAME, Schema.org, web semântica e linked data. Membro dos Grupos de Pesquisa Novas tecnologias em informação e Dados e metadados.

## Fernando César Lima Leite

Professor da Universidade de Brasília (Unb). Membro do corpo docente do Programa de Pós-Graduação em Ciência da Informação da UnB. Graduado em Biblioteconomia pela UnB. Mestre e Doutor em Ciência da Informação pela UnB. Atua nas linhas de pesquisa: ciência aberta (*open science*) em suas dimensões (social, política e tecnológica) e em estudos da informação no contexto de direitos humanos e minorias sociais.

## Flávia Maria Bastos

Coordenadora da Coordenadoria Geral de Bibliotecas (CGB) da Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP). Membro do corpo docente do Programa de Pós-Graduação em Planejamento e Gerenciamento de Sistemas de Informação da Faculdades Integradas Coração de Jesus (FAINC) e docente colaboradora do Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro (UNIRIO). Graduada em Biblioteconomia e Documentação pela UNESP. Mestre e Doutora em Ciência da Informação pela UNESP. Atua nas linhas de pesquisa: bibliotecas digitais, automação de bibliotecas, sistemas de gerenciamento de bibliotecas, bibliotecas colaborativas e repositórios institucionais.

## Gabriel David

Professor Associado da Faculdade de Engenharia da Universidade do Porto (FEUP). Doutor em Informática pela Universidade Nova de Lisboa (UNL). Atua nas linhas de pesquisa: gestão de informação, preservação digital e bases de dados. Membro do Grupo de Pesquisa InfoLab da FEUP.

## João Aguiar Castro

Graduado em História pela Universidade do Porto (U.Porto). Mestre em Ciência da Informação pela Faculdade de Engenharia da Universidade do Porto (FEUP). Doutorando em Media Digitais pela FEUP. Atua nas linhas de pesquisa: curadoria de dados, qualidade de metadados, gestão de dados de investigação. Membro do Grupo de Pesquisa InfoLab da FEUP.

## João Correia Lopes

Professor da Faculdade de Engenharia da Universidade do Porto (FEUP). Doutor em Ciência da Computação pela Universidade de Glasgow. Atua nas linhas de pesquisa: gestão de dados de investigação e infraestruturas para e-Ciência. Membro do Grupo de Pesquisa do Projeto TAIL.

## João Rocha da Silva

Professor convidado na Faculdade de Engenharia da Universidade do Porto (FEUP). Mestre em Engenharia Informática pela Universidade do Porto (U.Porto). Doutor em Engenharia Informática pela FEUP. Atua nas linhas de pesquisa: curadoria de dados, web semântica, *linked open data*, gestão de dados de investigação. Membro do Grupo de Pesquisa InfoLab da FEUP.

## Lucas Ângelo Silveira

Assistente de pesquisa no Instituto Brasileiro de Ciência e Tecnologia (IBICT). Graduado em Ciência da Computação pela Universidade Federal de Goiás (UFG). Mestre e Doutorando em Informática pela Universidade de Brasília (UnB).

## Lucas Rodrigues Costa

Assistente de pesquisa no Instituto Brasileiro de Ciência e Tecnologia (IBICT). Graduado em Computação pela Universidade de Brasília (UnB). Mestre e Doutorando em Informática pela UnB.

## Luis Fernando Sayão

Tecnologista da Comissão Nacional de Energia Nuclear (CNEN). Membro do corpo docente do Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro (UNIRIO) e do Programa de Pós-Graduação em Memória e Acervos da Fundação Casa de Rui Barbosa (FCRB). Graduado em Física pela Universidade Federal do Rio de Janeiro (UFRJ). Mestre e Doutor em Ciência da Informação pela UFRJ. Atua nas linhas de pesquisa: bibliotecas digitais, publicações eletrônicas, interoperabilidade, bases de dados, curadoria de dados de pesquisa e preservação digital.

## Marcos Galindo

Professor da Universidade Federal de Pernambuco (UFPE). Membro do corpo docente do Programa de Pós-Graduação em Ciência da Informação da UFPE. Graduado em Biblioteconomia pela UFPE. Mestre em História pela UFPE. Doutor em História pela Leiden University. Atua nas linhas de pesquisa: teoria da informação, memória e uso de tecnologia em sistemas memoriais.

## **Michelli Pereira da Costa**

Professora da Universidade de Brasília (UnB). Graduada em Biblioteconomia pela UnB. Mestra e Doutora em Ciências da Informação pela UnB. Atua nas linhas de pesquisa: dados de pesquisa, ciência aberta, acesso aberto, bibliotecas digitais e repositórios institucionais, política editorial, comunicação científica, revistas científicas.

## **Milton Shintaku**

Tecnólogo do Instituto Brasileiro de Ciência e Tecnologia (IBICT). Graduado em Ciências com Habilitação em Matemática pelo Centro Universitário de Brasília (UnICEUB). Mestre e Doutor em Ciências da Informação pela Universidade de Brasília (UnB). Atua nas linhas de pesquisa: repositório digital, ensino e aprendizagem da língua portuguesa como segunda língua a surdos usuários de Língua Brasileira de Sinais (LIBRAS) da Secretaria de educação do Distrito Federal (SEDF).

## **Ricardo Carvalho Amorim**

Consultor na área do desenvolvimento de aplicações móveis. Mestre em Engenharia Informática pela Faculdade de Engenharia da Universidade do Porto (FEUP). Atua nas linhas de pesquisa: aplicações móveis, cadernos de laboratório eletrônicos e gestão de dados de investigação. Membro do Grupo de Pesquisa InfoLab da FEUP.

## **Rúbia Gattelli**

Bibliotecária da Universidade Federal do Rio Grande (FURG). Graduada em Biblioteconomia pela FURG. Mestre em Ciência da Informação pela Universidade do Porto (U.Porto). Atua nas linhas de pesquisa: teoria da informação, organização e representação da informação e do conhecimento, recuperação da informação, tecnologias da informação e comunicação, sociedade da informação, gestão de dados científicos e curadoria digital.

## **Silvana Aparecida Borsetti Gregorio Vidotti**

Professora da Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP). Graduada em Matemática pela UNESP. Mestre em Ciência da Computação e Matemática Computacional pela Universidade de São Paulo (USP). Doutora em Educação pela UNESP. Atua nas linhas de pesquisa: ciência da informação com ênfase em tecnologias de informação e comunicação e em arquitetura da informação digital. Líder do Grupo de Pesquisa Novas Tecnologias em Informação. Bolsista de Produtividade em Pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Nível 2.



*Formato* 16 x 23 cm  
*Tipografia* PT Sans  
*Capa* Papel Triplex 240 g/m<sup>2</sup>  
*Miolo* Papel couché fosco 80 g/m<sup>2</sup>  
*Tiragem* 750 exemplares  
*Gráfica* Nome da Gráfica

Impresso no Brasil

*Printed in Brazil*

*Editora filiada a*



FSC®



A marca FSC® é a garantia de que a madeira utilizada na fabricação do papel deste livro provém de florestas que foram gerenciadas de maneira ambientalmente correta, socialmente justa e economicamente viável, além de outras fontes de origem controlada.





Considerando a relevância da reunião, do armazenamento, da disseminação e da preservação de dados e de informações digitais no contexto da produção científica de acesso aberto, Repositórios digitais: teoria e prática, enquanto produto derivado do I Fórum Nacional de Repositórios Digitais, realizado no ano de 2016, em Natal, Rio Grande do Norte, Brasil, apresenta discussões que apontam para as tendências teóricas e práticas na pesquisa e na implementação de repositórios digitais no cenário iberoamericano. Os capítulos, assinados por pesquisadores e profissionais que vêm dedicando seu trabalho ao Acesso Aberto, perpassam reflexões em relação às ferramentas e às tecnologias para gestão da informação, aos metadados, à curadoria digital, às políticas e às metodologias para implementação de repositórios digitais, trazendo contribuição significativa para os avanços dos estudos e práticas em repositórios digitais nos âmbitos científico, tecnológico e social.

Agência Brasileira do ISBN  
ISBN 978-85-7014-197-2



9 788570 141972